UNIVERSITÄT
BERN

Graduate School for Cellular and Biomedical Sciences

University of Bern

Improving the Clinical Use of Magnetic Resonance Spectroscopy for the Analysis of Brain Tumours using Machine Learning and Novel Post-Processing Methods

PhD thesis submitted by

Nuno Pedrosa de Barros

from Portugal

for the degree of

PhD in Biomedical Engineering

Supervisor: PD. Dr. Johannes Slotboom
Institute for Diagnostic and Interventional Neuroradiology
Faculty of Medicine of the University of Bern

Co-Advisor: Prof. Dr. Roland Kreis
Magnetic Resonance Spectroscopy and Methodology, Department of Clinical Research
Faculty of Medicine of the University of Bern

Copyright Notice

Abstract

Magnetic Resonance Spectroscopy (MRS) provides unique and clinically relevant information for the assessment of several diseases. However, using the currently available tools, MRS processing and analysis is time-consuming and requires profound expert knowledge. For these two reasons, MRS did not gain general acceptance as a mainstream diagnostic technique yet, and the currently available clinical tools have seen little progress during the past years.

MRS provides localized chemical information non-invasively, making it a valuable technique for the assessment of various diseases and conditions, namely brain, prostate and breast cancer, and metabolic diseases affecting the brain. In brain cancer, MRS is normally used for: (1.) differentiation between tumors and non-cancerous lesions, (2.) tumor typing and grading, (3.) differentiation between tumor-progression and radiation necrosis, and (4.) identification of tumor infiltration. Despite the value of MRS for these tasks, susceptibility differences associated with tissue-bone and tissue-air interfaces, as well as with the presence of post-operative paramagnetic particles, affect the quality of brain MR spectra and consequently reduce their clinical value. Therefore, the proper quality management of MRS acquisition and processing is essential to achieve unambiguous and reproducible results. In this thesis, special emphasis was placed on this topic.

This thesis addresses some of the major problems that limit the use of MRS in brain tumors and focuses on the use of machine learning for the automation of the MRS processing pipeline and for assisting the interpretation of MRS data. Three main topics were investigated: (1.) automatic quality control of MRS data, (2.) identification of spectroscopic patterns characteristic of different tissue-types in brain tumors, and (3.) development of a new approach for the detection of tumor-related changes in GBM using MRSI data.

The first topic tackles the problem of MR spectra being frequently affected by signal artifacts that obscure their clinical information content. Manual identification of these artifacts is subjective and is only practically feasible for single-voxel acquisitions and in case the user has an extensive experience with MRS. Therefore, the automatic distinction between data of good or bad quality is an essential step for the automation of MRS processing and routine reporting.

The second topic addresses the difficulties that arise while interpreting MRS results: the interpretation requires expert knowledge, which is not available at every site. Consequently, the development of methods that enable the easy comparison of new spectra with known spectroscopic patterns is of utmost importance for clinical applications of MRS.

The third and last topic focuses on the use of MRSI information for the detection of tumor-related effects in the periphery of brain tumors. Several research groups have shown that MRSI information enables the detection of tumor infiltration in regions where structural MRI appears normal. However, many of the approaches described in the literature make use of only a very limited amount of the total information contained in each MR spectrum. Thus, a better way to exploit MRSI information should enable an improvement in the detection of tumor borders, and consequently improve the treatment of brain tumor patients.

The development of the methods described was made possible by a novel software tool for the combined processing of MRS and MRI: SpectrIm. This tool, which is currently distributed as part of the jMRUI software suite (`www.jmrui.eu`), is ubiquitous to all of the different methods presented and was one of the main outputs of the doctoral work.

Overall, this thesis presents different methods that, when combined, enable the full automation of MRS processing and assist the analysis of MRS data in brain tumors. By allowing clinical users to obtain more information from MRS with less effort, this thesis contributes to

the transformation of MRS into an important clinical tool that may be available whenever its information is of relevance for patient management.

Keywords: Magnetic Resonance Spectroscopy (MRS), Magnetic Resonance Spectroscopic Imaging (MRSI), quality control, nosologic imaging, machine learning, brain tumors, glioblastoma multiforme, brain tumor delineation

" A day will come when war will seem as absurd between Paris and London, between Petersburg and Berlin, between Vienna and Turin, as it would today between Rouen and Amiens or between Boston and Philadelphia. A day will come when you – France, Russia, Italy, England, Germany – all you nations of the continent will merge, without losing your distinct qualities and your glorious individuality, in a close and higher unity to form a European brotherhood...

A day will come when the only fields of battle will be markets opening up to trade and minds opening up to ideas. A day will come when bullets and bombs will be replaced by votes, by universal suffrage of the peoples, by the venerable arbitration of a great sovereign senate which will be to Europe what Parliament is to England, (...) what the Legislative Assembly is to France. A day will come when we will display cannon in museums just as we display instruments of torture today, amazed that such things could ever have been. (...)

A day will come when we shall see those two immense groups, the United States of America and the United States of Europe, stretching out their hands across the sea, exchanging their products, their arts, their works of genius (...) And to bring about that day will not take another 400 years, for we are living in a fast-moving age. "

Victor Hugo, International Peace Congress, 1849.

Acknowledgments

Like anything in life, this thesis had a context that made it possible, a context created by the work of many others to whom I am deeply grateful. The work here presented results from a European project called TRANSACT (TRANSforming magnetic resonance spectroscopy into A Clinical Tool) that provided funding and allowed me to learn from several worldwide experts on the field of Magnetic Resonance Spectroscopy to whom I want to express my gratitude. Several years ago, my interest in Medical Image Processing started in Portugal, while studying Biomedical Engineering at Instituto Superior Técnico. During that period, I had the opportunity to study one semester at the Université Libre de Bruxelles (ULB) thanks to the Erasmus program. That European experience was crucial to deepen my knowledge and grow my interest in medical imaging. Without all the people that dedicated their lives to the European dream, this thesis wouldn't have existed, and it is to them that I dedicate the work here presented.

My Ph.D. "adventure" started officially on the 17th of December 2012, after sending an email to Dr. Johannes Slotboom entitled "Application to Brain Tumour / MRI Analyzer position (ESR8)". A few days after, I couldn't be happier when I received an answer saying that I was invited for an interview. A month passed and, on the 21st of January, I finally met Dr. Slotboom in person at the entrance of Inselspital, in Bern. The interview started with a walk through Inselspital's tunnel labyrinth, something that reminded of those movies where a victim is transferred to a secret place with his/her eyes covered. It was clear to me, even if I would get the job I would never be able to find that place again. I had a great time that day and, surprisingly, talking with Dr. Slotboom felt like talking to a friend. As everyone can guess now, I ended being the selected candidate. On the 30th of July 2013 I moved to Bern, and soon after Dr. Slotboom – my Ph.D. supervisor – became Hans. Learning from Hans' unique experience on the many different aspects of MRS (hardware, acquisition, data processing, algorithm development, clinical use, etc.), was an amazing opportunity for me. During the whole Ph.D., Hans was always available to discuss the latest results and plan the next steps. His contagious and true interest was essential to keep me motivated every time. Many of the discussions were done in the hospital, but many others were done during our traditional "beer-pizza" dinners. I thank you, Hans, for all the time that you dedicated as my supervisor, for the huge opportunity that you gave me to work on this amazing topic, for the freedom that you gave me to develop my own ideas, and for your friendship. It was a real honor for me to have you as my supervisor and to work with you.

Besides my supervisor, I want to thank the whole team of the Institute for Diagnostic and Interventional Neuroradiology (DIN) and of the Support Center for Advanced Neuroimaging (SCAN), who supported me always and gave me the perfect conditions to conduct my research. The team from the DIN is unique in their dedication to always keep innovating and improving medical practices. I especially want to thank Prof. Roland Wiest, for leading the SCAN group. Thank you, Roland, for your full support throughout my thesis, and for your unique and inspiring interest in new technologies to push forward medical imaging. I also cannot forget that the conditions that I had to work were only made possible by Prof. Gerhard Schroth, whose vision of a multidisciplinary team many years ago made the Institute what it is today, and Prof. Jan Gralla for currently leading this amazing team. Thank you, both, for everything. I would also like to thank Richard McKinley for all our discussions on machine learning; Urspeter Knecht for showing me MRS through the eyes of an MD; Martin Zbinden for all the uncountable hours that he spends to ensure that the SCAN computers are live and running; Manuela Pastore-Wapp for her complete dedication to the SCAN team; Christian Rummel for

all our interesting discussions; Claus Kiefer for his support with pulse sequence programming; and Samuel Stettler, for the time dedicated to the selection profile correction project. I shouldn't forget the brilliant work done by Jonathan Giezendanner and Manuel Schmid that helped me with SpectrIm, and by Martin Pletscher, my first Master student, who did an amazing work on clustering MRSI data of GBM patients. Thank you!

Also from Bern, I want to thank my co-supervisor Prof. Roland Kreis, for his guidance and support during my thesis. Roland was always available to discuss my progress during the Ph.D. and provided invaluable insights on the different aspects that are currently regarded as limitations for the clinical use of MRS. This thesis was shaped, not only by our discussions but also by several of his publications. Thank you, Roland.

From the same group, I would also like to thank Pruthvi, for all the nice time spent together in Bern and in every ISMRM meeting, as well as for all our discussions, many of them in the train, while returning from lectures at the ETH, in Zürich.

Moving to Spain, Barcelona, I want to thank the people from the Universitat Autònoma de Barcelona, namely Margarida Julià-Sapé, Prof. Carles Arús and Victor Mocioiu. I learned a lot from them, namely in the use of convex-Non-negative Matrix Factorization (cNMF) to assist the interpretation of MRSI data. Besides our work together, I couldn't have been received better in Barcelona. I also remember the nice time spent with Victor during his visit to Bern, namely during our jogging sessions through the forest.

During the whole project, I counted with full support from the jMRUI team from the Czech Republic: Jana Starčuková and Michał Jabłoński. Michał also had the opportunity to come to Bern, where we laughed with his stories of the "bread baked by the king". I remember with joy our hike and the treasure hunt through the Gurten. Thank you, Jana and Michał.

The story of this thesis would've not been complete without the time spent at Icometrix in Leuven, Belgium. There I was received by an amazing team and I learned different aspects of image co-registration. I specifically want to thank Diana Sima, Annemie Ribbens, Dirk Smeets, Dirk Loeckx and Saurabh Jain. I also want to thank Anke Maertens, without whom I would never have met Kaatje.

Kaat was a very lucky finding during my stay in Leuven, in the beginning of the Ph.D. program. I thank you, Kaat, for all the crazy facetime performances that make me laugh every time, and for all the love and joy that kept me going throughout my doctoral work. You gave a new meaning to my Ph.D. experience and I look forward to our future together.

While everything was happening throughout Europe, many of the most important people for me were in Lisbon. Without their support, I would have never been writing these lines. I specifically want to thank my parents, my brother André and his wife Mariana, my sister Joana, my grandma Belmirinha, and my grandparents Lila, Eloi and Z'Alberto. I also want to thank my aunts João e Alda, my uncle Titó, and my dear cousins António Gonçalo, António, Inês, Matilde e Leo. I still remember as if it was yesterday the dinner at Carvoeiro, the day before stepping on the plane that brought me to Bern. You know what you mean to me and I thank you for all your support during all moments of my life. I also want to thank all my friends from Lisbon, especially those that, with their visits to Switzerland, shortened the distance between Bern and Lisbon: Xico, Carlos, Pedro, Jojo (doroteias), Anta, Conchita, Diana, Jojo (biomed), Nuno, Manel, Zeca, Miguel, Susana, Ana, Andreia, André,...

Finally, I want to thank all the amazing people that I met during the last years, who made this Ph.D. such a nice experience. I can not help to mention: the friends from PH8, Megha, Vijay, Vega, Anil, Anna-Lena, Francesco, Paula, Alessandro, Christian, Mathias, Chantal, Paulo, Anna; the Bernese gang, Tigran, Gerrit, Michiel, Joh; the friends from the "turtle club", Victor,

José, Pedro, Magda, Heli, Toni, Kevin, Marta; and last but not least, all my TRANSACT friends: Pruthvi, Michal, Victor, Miguel, Veronika, Nassim, Saurabh, Adrian, Sana, Akila, Claudio, Ross and Iveta.

Nuno, Bern, 2017.

List of Abbreviations

ADC Apparent Diffusion Coefficient.
ADC Analog to Digital Converter.
ARSOS All Rank Selected Order Statistic.
AUC Area Under the Curve.
BraTSIS Brain Tumor Spectroscopic Imaging Segmentation.
BraTumIA Brain Tumor Image Analysis.
Cho Choline.
cNMF convex-Non-negative Matrix Factorization.
Cr Creatine.
CRLB Cramér-Rao Lower Bounds.
CSA Chemical-Shift Artefact.
CSF Cerebral Spinal Fluid.
CSI Chemical Shift Imaging.
EDT Expected Distance to Tumor.
EPI Echo Planar Imaging.
EPSI Echo Planar Spectroscopy Imaging.
ER-ARSOS Error Estimation - All Rank Selected Order Statistic.
E-Tumor Enhancing Tumor.
FD Frequency-Domain.
FLAIR Fluid Attenuated Inversion Recovery.
FWHM Full-Width at Half Maximum.
GABA Gamma-AminoButyric Acid.
GBM GlioBlastoma Multiforme.
Gln Glutamine.
Glu Glutamate.
Glx Glutamine + Glutamate.
GM Grey-Matter.
HLSVD Hankel-Lanczos Singular Value Decomposition.
Lac Lactate.
LCModel Linear Combination of Model spectra.
Lip Lip1.3+Lip0.9.
Lip0.9 0.9 ppm signals originating from the -CH3 moiety of Lipid molecules (triglycerides).
Lip1.3 1.3 ppm signals originating from the -CH2 moiety of Lipid molecules (triglycerides).
LOOCV Leave-One-Out Cross-Validation.
MEGA-PRESS MEscher-GArwood - Point Resolved Spectroscopy.
MR Magnetic Resonance.
MRI Magnetic Resonance Imaging.
MRS Magnetic Resonance Spectroscopy.

MRS-FSD Magnetic Resonance Spectroscopy - Feature Spatial Distribution.
MRSI Magnetic Resonance Spectroscopic Imaging.
NAA N- Acetyl Aspartate.
NE-Tumor Non-Enhancing Tumor.
NMF Non-negative Matrix Factorization.
NMR Nuclear Magnetic Resonance.
PRESS Point Resolved Spectroscopy.
PSF Point Spread Function.
QUALITY QUAntification improvement by converting LIneshapes to the lorentzian TYpe.
QUEST QUantitation based on quantum ESTimation.
RF Random Forests.
RF Radio Frequency.
SADLOVE Single-shot ADiabatic LOcalized Volume Excitation.
SAR Specific Absorption Rate.
SE Spin Echo.
SESI Spin Echo Spectroscopic Imaging.
sMRI structural Magnetic Resonance Imaging.
SNR Signal to Noise Ratio.
SPICE SPectroscopic Imaging by exploiting spatiospectral CorrElation.
STEAM Stimulated Echo Acquisition Mode.
STO1 Spectroscopy Test Object 1.
STO2 Spectroscopy Test Object 2.
STV Solid Tumor Volume.
SVS Single Volume Spectroscopy.
TD Time-Domain.
TE Echo Time.
TR Repetition Time.
VOI Volume Of Interest.
WHO World Health Organization.
WM White-Matter.

Contents

1 | Introduction

Magnetic Resonance Spectroscopy (MRS) enables the non-invasive retrieval of localized biochemical information from different parts of the human body and allows to quantify different metabolites that can not be analyzed using conventional MRI techniques. The ability to detect metabolic changes in various tissues makes it a very appealing technique for initial diagnostics, treatment planning, and follow-up in many diseases and conditions.

Currently, the main clinical applications of MRS are found in oncologic imaging [García-Figueiras 2016] of the brain [Preul 1996, Howe 2003, Dowling 2001, Majós 2004], prostate [Costello 1999, Mueller-Lisse 2007, Tayari 2017] and breast [Suppiah 2013, Cen 2014, Baltzer 2013], as well as in the evaluation of several metabolic diseases affecting the brain [Zimmerman 2011, Harting 2015, Razek 2014]. Other interesting applications include the assessment of neonatal hypoxic injury [Boichot 2006, Peden 1993] and the measurement of intramyocellular lipid content in skeletal muscle to assess insulin sensitivity in pre-diabetic patients [Boesch 1999, Boesch 2006].

Proton MRS, the most common type of MR spectroscopy, can be performed in almost every MRI scanner if the right software (i.e. pulse sequence) is installed. Despite this and of its relevance, MRS is clinically used only in a few specialized centers.

1.1 Current limitations and challenges in clinical MRS

In 2012, a survey conducted by R. Kreis [Kreis 2012], aimed at collecting information regarding the reasons that limit the use of MRS in clinical practice. In that survey, 35 different professionals from various hospitals of more than 15 countries, were asked the following question:

What does it take for an MR site to perform MRS exams that can be useful for patient management?.

Figure 1.1 shows the different items that could be selected, as well as the percentage of participants that selected each one of them. In the same figure, "orange" was used to highlight the topics from the list that, directly or indirectly, were addressed in this thesis. The results of the survey show that the need for expertise in the different aspects associated with MRS acquisition, post-processing, and interpretation, is regarded as the most limiting factor for the clinical use of MRS. This thesis presents methods that aim at addressing this problem by (1.) enabling the fully automatic post-processing of MRS, and (2.) assisting clinicians in the analysis and interpretation of MRS data. Moreover, a complete chapter was dedicated to the topic of quality management in *in vivo* MRS, whose content should assist non-experienced users with the different quality-related technological aspects that are essential to guarantee unambiguous and reproducible results.

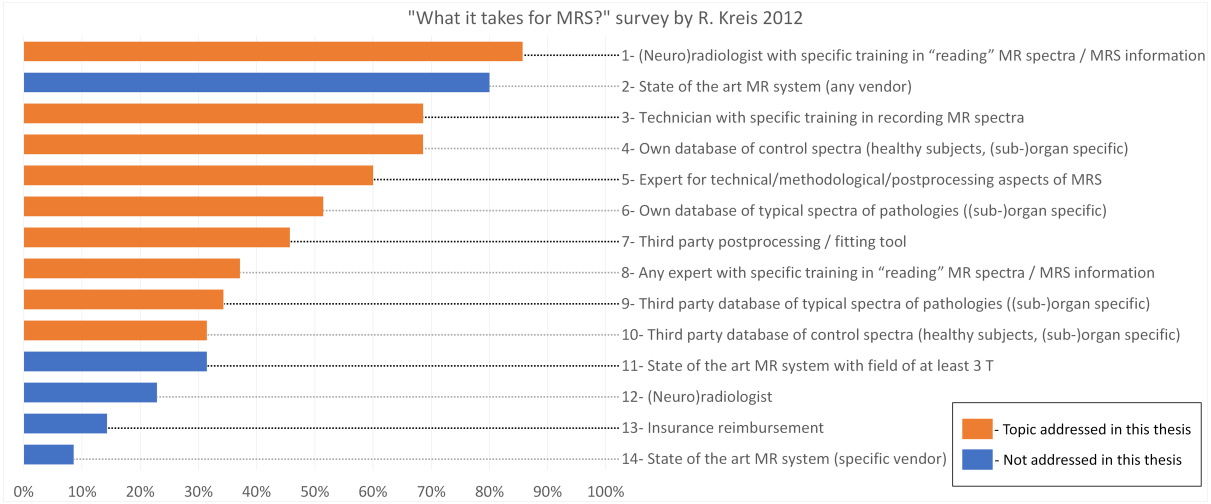

Figure 1.1: Results of a survey conducted by R. Kreis in 2012 [Kreis 2012] where the participants were asked the following question: "What does it take for an MR site to perform MRS exams that can be useful for patient management?". Each participant could select any number of items from the list shown in the figure. The graph shows the percentage of participants that selected each item. From those topics, the ones that are, directly or indirectly, addressed in this thesis are marked in orange, and the remaining ones in blue.

1.2 Hypothesis and aim of the thesis

As discussed in the previous section, the need for expertise is one of the main limiting factors for the clinical use of MRS. Therefore, new methods are needed that simplify the use of MRS and allow clinical centers with less experience to use the technique whenever its information may be of relevance for patient management. The need for the simplification of the analysis is further emphasized in a review paper published in 2016 on the use of MRS in oncologic imaging [García-Figueiras 2016]:

"[1]*H-MRS has achieved great strides as a molecular imaging technique since its introduction, and its scope in many clinical scenarios and research settings is rising. However, MRS needs expertise and is time-consuming, which limit its clinical applicability. In this setting, spectra analysis needs to be simplified.*"

This thesis addresses the issues related to the need for expertise and to the time required for the processing and analysis of MRS. More precisely, the work here presented aimed at improving and fully automatizing MRS post-processing, as well as assist the analysis and interpretation of MRS data. This was done considering the following hypothesis:

Using machine learning methods, it is possible to fully automatize MRS post-processing, as well as produce results that may be easily interpreted by non-experienced clinical MRS users.

The focus on the automation and simplification of MRS is a constant aspect of the topics presented in this thesis.

1.3 Thesis structure

This thesis is organized as follows:

- **Part I: Theoretical Background** - covers the theory that underlies the topics presented in this thesis.

  - **Chapter 2: From NMR to MRI and MRS** - introduces the reader to the main theoretical concepts of Nuclear Magnetic Resonance (NMR), Magnetic Resonance Imaging (MRI) and Magnetic Resonance Spectroscopy (MRS).
  - **Chapter 3: MRS Data Processing & Machine Learning** - covers the main post-processing steps used in MRS.
  - **Chapter 4: Brain Tumor Characterization using *in vivo* proton MRS** - presents the main metabolites visible in *in vivo* proton MR spectra of the brain and discusses the main applications of MRS for the characterization of brain tumors.

- **Part II: Quality management of MRS and automation of quality control** - covers different quality related aspects of MRS.

  - **Chapter 5: Quality Management in *in vivo* proton MRS** - discusses various aspects related to the quality management of *in vivo* proton MRS that are essential to guarantee unambiguous and reproducible results with MRS.
  - **Chapter 6: Automatic Quality Control of proton MRSI data** - presents a machine learning based method for the automatic quality control of MRS data that was trained to mimic the judgment of MRS experts on the acceptance or rejection of spectra based on the subjective quality of the MRS signal.
  - **Chapter 7: Improving labeling efficiency in automatic QC of MRSI data** - continues the work of the previous chapter and proposes a new approach to improve the efficiency of labeling the examples used to train quality control classifiers.

- **Part III: Assisting the interpretation of brain tumor MRS data** - covers different methods to improve and simplify the analysis and interpretation of brain tumor MRS data.

  - **Chapter 8: Highlighting differences between GBM and metastasis using cNMF** - presents a method based on convex non-Negative Matrix Factorization (cNMF) that was used to highlight differences between GBM and Metastasis, which are two of the most common types of brain tumors and for which the correct differentiation using conventional medical imaging/spectroscopy is not always straightforward.
  - **Chapter 9: Automatic tissue-type classification of ^{1}H-MRSI spectra in GBM patients** - presents the results of an approach where the MRSI data of GBM patients is clustered and the different prototype spectra of each cluster are then used to classify new spectra.
  - **Chapter 10: BraTSIS: Brain Tumour Spectroscopy Imaging Segmentation protocol** - continues the work of the previous chapter and presents a decision-tree based algorithm that allows to consistently identify the main types of tissue-type spectroscopic patterns seen in MRSI examinations of GBM patients.
  - **Chapter 11: MRS features and distance to MRI-visible Solid Tumor in GBM** - explores the relationship between the distance to solid tumor as is visible in structural MR-imaging, and MR-spectroscopy derived features. By doing so, the spatial properties of different metabolite features are compared. Moreover, a new method for the analysis of MRSI data of brain tumors is presented that may improve the detection of tumor-related changes in brain MRSI data. This method predicts the distance to the solid tumor for each voxel based solely on its metabolite content allowing to create Expected Distance to Tumor (EDT) maps.

- The thesis finishes with **Chapter 12 (Overall Discussion & Outlook)**, where the main results are discussed and a few possible directions to continue the work here presented are indicated.

Part I

Theoretical Background

2 | From Nuclear Magnetic Resonance to Magnetic Resonance Imaging and Spectroscopy

2.1 Nuclear Magnetic Resonance (NMR)

The Nuclear Magnetic Resonance (NMR) effect was first observed by Isidor Isaac Rabi in 1938 [Rabi 1938], a discovery that justified his recognition with a Nobel prize in physics in 1944 *"for his resonance method for recording the magnetic properties of atomic nuclei"*. Cornelius Jacobus Gorter from the Zeeman laboratory in Amsterdam attempted to measure the NMR effect a few years before [Gorter 1942] but his experiments were unsuccessful. Besides Gorter's lack of success, he visited Rabi's lab in September 1937, an event that is believed to have been crucial for Rabi's results. The initial discovery was made using a beam of lithium chloride molecules in a vacuum [Rabi 1938], and it was not until 1945 that NMR was shown in condensed matter by Felix Bloch and Edward Purcell working independently [Bloch 1946, Purcell 1946]. Bloch and Purcell received the physics Nobel prize in 1952 *"for their development of new methods for nuclear magnetic precision measurements and discoveries in connection therewith"*.

The discovery of NMR led to the invention of two technologies that revolutionized chemistry and medicine and impacted many other areas: NMR spectroscopy and Magnetic Resonance Imaging (MRI). Besides exploring the exact same principles, NMR spectroscopy and MRI became two distinct fields of research and, today, are normally used in different contexts. On one hand, MRI is an essential tool used in modern medicine that enables the non-invasive *in-vivo* visualization of the structure and function of different organs and tissues. NMR spectroscopy, on the other hand, has become an important tool used in chemistry and physics for analyzing the chemical composition and properties of both organic and inorganic samples. Magnetic Resonance Spectroscopy (MRS), the technique explored in this thesis, is found at the intersection between MRI and NMR spectroscopy. MRS refers to the use of NMR spectroscopy for the analysis of *in vivo* samples and is normally performed using the same hardware as in MRI.

In the next sections, the NMR principles will be presented, followed by the basic concepts that underlie MRI and MRS.

2.2 NMR principles

2.2.1 Nuclear spin and magnetic momentum

Nuclear Magnetic Resonance is a physical process where a nucleus absorbs or emits a photon whose energy is in resonance with the transition between different nuclear spin angular momentum energy levels. The effect only occurs for nuclei with a non-zero spin quantum number (I) that are subject to an external magnetic field. In the absence of an external magnetic field, and for nuclei with I equal to zero, there is only a single energy level and, consequently, nuclear

Table 2.1: Isotope natural abundance, spin quantum number, number of protons and neutrons and gyromagnetic ratios for different nuclei. Source: Department of Chemistry Instrumentation Facility, MIT, Massachusetts, USA, 2017

Isotope	**Abundance (%)**	I	**#protons**	**#neutrons**	γ**(MHz/T)**
^{1}H	99.98	1/2	1	0	42.58
^{2}H	1.5×10^{-2}	1	1	1	0.15
^{3}He	1.3×10^{-4}	1/2	2	1	0.71
^{13}C	1.11	1/2	6	7	0.33
^{14}N	99.63	1	7	7	0.29
^{15}N	0.37	1/2	7	8	1.40
^{17}O	3.7×10^{-2}	5/2	8	9	1.34
^{19}F	100	1/2	9	10	6.94
^{23}Na	100	3/2	11	12	0.28
^{31}P	100	1/2	15	16	1.53

magnetic resonance is not possible. The spin quantum number is a property of each nucleus that depends on the number of protons and neutrons:

- For nuclei with odd number of protons and odd number of neutrons, $\boldsymbol{I}$ is a positive integer ($I = \{1, 2, 3, ...\}$);
- For nuclei with odd number of protons and even number of neutrons, or vice versa, $\boldsymbol{I}$ is a positive half-integer ($I = \{1/2, 3/2, 5/2, ...\}$);
- For nuclei with even number of protons and even number of neutrons, $\boldsymbol{I}$ is zero ($I = 0$);

The spin quantum number of a nucleus determines the quantization of the nuclear angular momentum (J). The magnitude of J is given by:

$$|J| = \hbar\sqrt{I(I+1)} \tag{2.1}$$

where $\hbar$ is the reduced Planck constant ($1.055\times10^{-34} Js/rad$). Because nuclei contain charged particles (protons), its rotation is associated with a magnetic momentum (μ). For a given nucleus, the relation between its magnetic momentum (μ) and angular momentum (J) is given by a physical constant named *gyromagnetic ratio* (γ):

$$\gamma = \mu/J \tag{2.2}$$

Table 2.1 shows the spin quantum number and gyromagnetic ratios for various nuclei corresponding to different isotopes.

In the presence of a strong magnetic field (B_0) in a given spatial direction z, the magnetic momentum vector in the same direction can only have the following values:

$$\mu_z = \gamma\hbar m_I, \qquad m_I = -I, -I+1, ..., I-1, I \tag{2.3}$$

where m_I is the nuclear magnetic quantum number. The x and y components of the vector can assume any value, provided that the magnitude of μ respects equation 2.1. This means that, in the case of the proton, due to its spin quantum number of 1/2, the magnetic momentum vector makes an angle of approximately 54.74° with its z component (see Figure 2.1):

$$\theta = arcos(\frac{|\mu_z|}{|\mu|}) = arcos(\frac{\gamma\hbar|m_I|}{\gamma\hbar\sqrt{I(I+1)}}) \overset{I=\frac{1}{2}}{=\!=} arcos(\frac{1}{\sqrt{3}}) \approx 54.74° \tag{2.4}$$

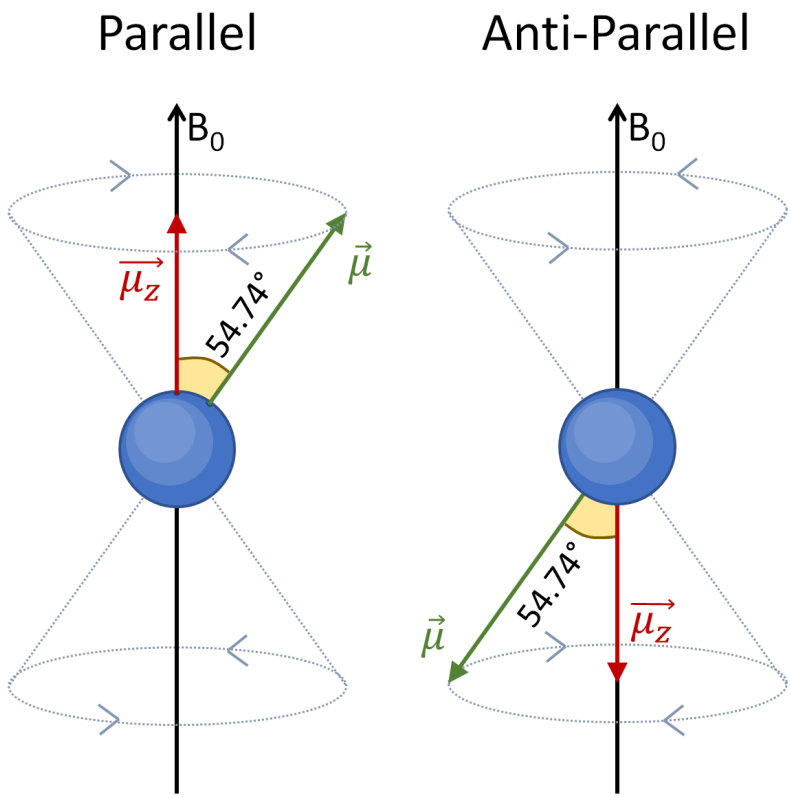

Figure 2.1: Possible orientations for a nucleus with a spin quantum number of 1/2. The figures show the precession movement of the protons as a result of the torque exerted by the static magnetic field B_0.

2.2.2 Larmor frequency

The energy associated with the magnetic momentum depends on the strength of the external magnetic field (B_0) and the magnetic momentum in the direction of the field (μ_z):

$$E = -\mu_z B_0 = -\gamma \hbar m_I B_0, \qquad m_I = -I, -I+1, ..., I-1, I \tag{2.5}$$

We will now focus on isotopes for which $I = 1/2$, such as ^1H. For these nuclei, there are only 2 energy levels:

$$E = \pm \frac{1}{2} \gamma \hbar B_0 \tag{2.6}$$

The two energy levels correspond to the magnetic momentum in the z direction being either positive or negative, meaning that the magnetic momentum vector is either *parallel* or *anti-parallel* to the static external magnetic field (B_0). The energy associated with the transition between these two levels is given by:

$$\Delta E = \gamma \hbar B_0 \tag{2.7}$$

In order to calculate the corresponding resonance frequency, we need to consider the relation between the energy of a photon and its frequency (Planck-Einstein relation):

$$E = \hbar \omega \tag{2.8}$$

Consequently, the resonance frequency, or *Larmor frequency* (ω_0), is given by:

$$\omega_0 = \gamma B_0 \tag{2.9}$$

2.2.3 Population difference between orientations

In an NMR experiment, the detected signal corresponds not to a single nucleus but to the combined effect of all nuclei of the sample. The NMR signal is proportional to the population difference between the *parallel* and *anti-parallel* states. The ratio between the two populations can be calculated considering the Boltzmann equation:

$$\frac{N_{antiparallel}}{N_{parallel}} = e^{-\frac{\Delta E}{kT}} = e^{-\frac{\gamma\hbar B_0}{kT}} \approx 1 - \frac{\gamma\hbar B_0}{kT} \tag{2.10}$$

where k is the Boltzmann constant ($1.381 \times 10^{-23} J/K$) and T is the temperature in Kelvin. Following the first-order approximation of equation 2.10, the population difference is given by:

$$\Delta N = N_{parallel} - N_{antiparallel} \approx N\frac{\gamma\hbar B_0}{2kT} \tag{2.11}$$

where N is the total number of nuclei in the sample. This equation shows that in the case of ^1H, at room temperature (296.15K), and for a magnetic field of 1.5T, there is only a difference of approximately 5 nuclei between the two states for every million nuclei.

2.2.4 Macroscopic Magnetic Dipolar Momentum

The net magnetization of a sample ($\vec{M_0}$) corresponds to the combined effect of all nuclei present in a given volume. Therefore, in equilibrium, it is given by:

$$\vec{M_0} = \frac{\sum \vec{\mu}}{V} = \frac{\gamma\hbar\Delta N}{2V}\Delta N\vec{e_z} \approx \frac{N\gamma^2\hbar^2 B_0}{4kTV}\vec{e_z} \tag{2.12}$$

where V is the volume of the sample. From this expression, it is seen that, in equilibrium, $\vec{M_0}$ has only a component in the z-direction. This results from the fact that each individual magnetic momentum vector assumes a random orientation in the x and y directions, which means that all vector components in the plane perpendicular to the external magnetic field cancel each other.

2.2.5 Bloch Equations and Relaxation

In every NMR experiment, the information from our sample is collected by first interacting with the spins of the sample in such a way that these will respond differently depending on the properties that are targeted by the experiment. The main interaction with the different nuclei is made by applying a time-varying magnetic field perpendicular to the main static magnetic field B_0, that superimposes B_0. Consequently, the total magnetic field $\vec{B}(t)$ is given by:

$$\vec{B}(t) = (B_x(t), B_y(t), B_0) \tag{2.13}$$

The evolution of $\vec{M}$ as a result of $\vec{B}$ can be modeled by the Bloch equations:

$$\begin{aligned}
\frac{dM_x}{dt} &= \gamma(M_y B_0 - M_z B_y) - \frac{1}{T_2}M_x \\
\frac{dM_y}{dt} &= \gamma(M_z B_x - M_x B_0) - \frac{1}{T_2}M_y \\
\frac{dM_z}{dt} &= \gamma(M_x B_y - M_y B_x) - \frac{1}{T_1}(M_z - M_0)
\end{aligned} \tag{2.14}$$

Table 2.2: Relaxation times for different tissues at 1.5T. Sources: [Webb 2003, Meier 2012]

Tissue	T_1 **(ms)**	T_2 **(ms)**
Fat	260	80
Muscle	870	45
Gray Matter	900	100
White Matter	780	90
Cerebrospinal Fluid	2400	160
Liver	500	40
Kidney	650	450

where M_z is referred to as *longitudinal magnetization*, and M_x and M_y as *transverse magnetization*. In these equations it is possible to distinguish two main effects:

1. The *torque* that the magnetic field $\vec{B}$ exerts on $\vec{M}$;

2. The *relaxation* of the magnetization with time-constants T_1 and T_2. These constants depend on the local environment and change between different tissues.

The two time-constants that describe the relaxation of $\vec{M}$, T_1 and T_2, are manifestations of different effects. The first constant, T_1, is a result of the loss of energy of the nuclei to the surrounding atoms, bringing the system back to the equilibrium. Consequently, T_1 is called *spin-lattice relaxation time*. In contrast, T_2 does not depict the exchange of energy but the loss of phase coherence between the different magnetic momentum vectors of the multiple nuclei. This loss of coherence leads to the cancellation of the transverse components of $\vec{M}$. Given that this effect is related to the phase-coherence between different spins, T_2 is called *spin-spin relaxation time*. It is important to note that two different phenomena are responsible for the loss of phase coherence between spins. One one hand, the neighborhood of each spin and the interactions with the surrounding atoms causes each nuclei to experience a slightly different magnetic field that leads to the accumulation of a phase difference between spins. This mechanism for loosing phase-coherence is called the "true" T_2 relaxation. Additionally to this effect, the fluctuations in the magnetic field due to hardware imperfections (e.g. imperfect shimming), as well as magnetic susceptibilities differences between different materials, cause local field inhomogeneities that accelerate the loss of coherence between spins. Therefore, the observed relaxation time of the transverse magnetization, denoted (T_2^*), is a combination of these two effects:

$$\frac{1}{T_2^*} = \frac{1}{T_2^+} + \frac{1}{T_2} \tag{2.15}$$

where T_2^+ represents the contribution of the local field inhomogeneities to T_2^*. For the sake of accuracy, it should be clarified that in the Bloch equations above presented, the constant T_2 is, in fact, T_2^*.

The different time-constants, T_1, T_2 and T_2^*, can vary considerably between different tissues. This allows MRI to have a high soft-tissue contrast, which is one of the main advantages of MRI relatively to other medical imaging techniques such as computer-tomography (CT) or ultrasound (US). Table 2.2 shows the observed approximate relaxation times (T_1 and T_2) for different tissues at 1.5T.

2.2.6 NMR Signal

As it was shown in the Bloch equations, the static magnetic field $\vec{B_0}$ exerts a torque on $\vec{M}$. In the presence of transverse magnetization, the torque causes $\vec{M}$ to precess around the z-axis, creating a time-varying magnetic field that can be detected using a coil with a cross-section perpendicular to the transverse plane xOy. The relation between a time-varying magnetic flux and the electric field is described by the Maxwell-Faraday equation:

$$\nabla \times \vec{E} = -\frac{d\vec{B}}{dt} \tag{2.16}$$

The integral of the curl of the electric field over the cross-section of the coil corresponds to an electromotive force that creates a time-dependent voltage ($u(t)$) between the ends of the coil. The voltage $u(t)$ is proportional to the precession of the transverse components of $\vec{M}$. More specifically, if we consider that the magnetization is rotated 90° so that M_z is fully converted in M_x, and ignoring relaxation effects:

$$u(t) \propto \frac{dM_x}{dt} = j\omega_0 M_0 e^{j\omega_0 t} \approx jN\frac{\gamma^3\hbar^2 B_0^2}{4kT}e^{j\omega_0 t} \tag{2.17}$$

Equation 2.17 shows that the NMR signal scales quadratically with the magnetic field. However, this does not necessarily translate into a quadratic relation between magnetic field strength and the intrinsic signal-to-noise-ratio (SNR) of a given MR system, since the noise also increases with B_0. As described by [Ocali 1998], for low field strengths ($< 3T$), the relation between B_0 and SNR is approximately linear. However, for higher field strengths the relation between these two quantities becomes more complex due to the interaction between the used radio-frequency waves and the sample (e.g. human body) that create spatial dependencies between the SNR and B_0.

2.2.7 Spatial Localization and Small Tip-Angle Approximation

In *in vivo* NMR, it is essential to be able to selectively interact with spins of a given spatial region. In order to achieve this, magnetic field gradients can be used in combination with a time-varying magnetic field, i.e. an RF-pulse. The gradients superimpose $\vec{B_0}$, creating a spatially dependent resonance frequency ω_L:

$$B(x,y,z) = B_0 + G_y y + G_x x + G_z z \implies \omega_L(x,y,z) = \omega_0 + \gamma(G_y y + G_x x + G_z z) \tag{2.18}$$

where G_x, G_y and G_z are the linear gradients in the x, y and z directions. Once these gradients are active, an RF-pulse with a narrow band of frequencies will select only those regions where the resonance conditions are met.

In order to better understand the relation between the RF-pulse and how the transverse magnetization varies in a given direction, the Bloch equations will be solved for the particular case of the selection of a slice orthogonal to $\vec{B_0}$ using a linear gradient in the z-direction.

Before solving the equations, two concepts will be introduced that simplify its resolution. The first one corresponds to the concept of the *rotating frame of reference*, more precisely, a coordinate system that rotates at the Larmor frequency about the z-axis. Given that the effect of the static component of the magnetic field B_0 is a precession of $\vec{M_0}$ about the z-axis, in the rotating frame this movement is "eliminated", and consequently, considering $B_z = B_0 + G_{sel}z$ the effective magnetic field becomes:

$$\vec{B_{eff}} = B_{1,x'}(t)\vec{e_{x'}} + B_{1,y'}(t)\vec{e_{y'}} + (B_z - \frac{\omega_0}{\gamma})\vec{e_z} = B_{1,x'}(t)\vec{e_{x'}} + B_{1,y'}(t)\vec{e_{y'}} + G_{sel}z\vec{e_z} \tag{2.19}$$

where x' and y' correspond to the transverse coordinates in the rotating frame of reference (the z coordinate is identical for both coordinate systems). The second concept that will be introduced is the *small tip-angle approximation* which assumes that the pulse does not change the longitudinal magnetization, i.e. $M_z(t) \approx M_0$. Thus, if the relaxation components are ignored, the Bloch equations can be written as:

$$\begin{aligned} \frac{dM_{x'}}{dt} &= \gamma(M_{y'}G_{sel}z - M_0B_{1,y'}(t)) \\ \frac{dM_{y'}}{dt} &= \gamma(M_0B_{1,x'}(t) - M_{x'}G_{sel}z) \\ \frac{dM_z}{dt} &= 0 \end{aligned} \tag{2.20}$$

If the transverse magnetization $M_{xy}(t)$ is defined as the complex value

$$M_{xy}(t) = M_{x'}(t) + iM_{y'}(t) \tag{2.21}$$

with $i = \sqrt{-1}$, and the complex time-dependent field originated by the RF-pulse $B_1(t)$ as

$$B_1(t) = B_{1,x'}(t) + iB_{1,y'}(t) \tag{2.22}$$

the evolution of the transverse magnetization can be written as

$$\frac{dM_{xy}}{dt} = i\gamma M_0B_1(t) - i\gamma M_{xy}(t)G_{sel}z \tag{2.23}$$

As described in [Slotboom 1993] and [Noll 2006], the solution for the differential equation 2.23 for a pulse with duration τ is given by:

$$M_{xy}(z,\tau) = i\gamma M_0 e^{-i\gamma G_{sel}z\tau} \int_0^\tau B_1(t)e^{i\gamma G_{sel}zt}dt \tag{2.24}$$

To better understand the meaning of this equation, a variable substitution is performed, $t' = t - \tau/2$. Moreover, it is assumed that the RF is symmetrical around $\tau/2$ and is zero outside $[0,\tau]$:

$$\begin{aligned} M_{xy}(z,\tau) &= i\gamma M_0 e^{-i\gamma G_{sel}z\tau} \int_{-\tau/2}^{\tau/2} B_1(t'+\tau/2)e^{i\gamma G_{sel}z(t'+\tau/2)}dt' \\ &= i\gamma M_0 e^{-i\gamma G_{sel}z\tau/2} \int_{-\tau/2}^{\tau/2} B_1(t'+\tau/2)e^{i\gamma G_{sel}zt'}dt' \\ &= i\gamma M_0 e^{-i\gamma G_{sel}z\tau/2} \int_{-\infty}^{\infty} B_1(t'+\tau/2)e^{i2\pi\frac{\gamma G_{sel}}{2\pi}zt'}dt' \\ &= i\gamma M_0 e^{-i\gamma G_{sel}z\tau/2}\mathcal{F}^{-1}[B_1(t'+\tau/2)]|_{\omega'=\frac{\gamma G_{sel}}{2\pi}z} \end{aligned} \tag{2.25}$$

This result shows that the profile of the slice selected with the RF-pulse is proportional to its spectrum, with a conversion between frequency and z-coordinate given by:

$$\omega = \frac{\gamma G_{sel}}{2\pi}z \tag{2.26}$$

Moreover, the RF-pulse causes a phase-shift in the z-direction of:

$$\Delta\phi = e^{-i\gamma G_{sel}z\tau/2} \tag{2.27}$$

which can be easily eliminated if a gradient with slope $-G_{sel}$ is applied during half the duration of the pulse, i.e., $\tau/2$.

2.2.8 RF-pulses

As was shown in the previous subsection, RF-pulses can be used to convert longitudinal magnetization into measurable transverse magnetization. Such pulses are named *excitation pulses*. In general, RF-pulses can be divided into the following types ([Slotboom 1993]):

- *Non-self-refocused excitation pulses*. This corresponds to the RF-pulse type described in the previous section that aims at transforming longitudinal magnetization into transverse magnetization. As shown in the previous section (eq. 2.27), the selected region is left with a phase-shift that needs to be corrected afterward. Examples of such pulses include the Gauss and the sinc pulses.

- *Self-refocused excitation pulses*. Such pulses are a variation of the previous type that is engineered to eliminate the phase-shift created by the RF-pulse [Pauly 1991]. The advantage of these pulses is that immediate signal acquisition after the pulse can take place.

- *Inversion pulses*. Inversion pulses aim at inverting the longitudinal magnetization. Such pulses are the basis of sequences such as Inversion Recovery sequences, where the differences in recovery times between different tissues or metabolites are used to selectively eliminate the signal of one of the sources. An example of such a sequence is FLAIR (Fluid Attenuated Inversion Recovery), where the signal corresponding to the CSF is selectively eliminated. In MRS, the same technique can be used to separate signals of macromolecules, i.e. lipids and proteins, from metabolites with longer T_1 relaxation times (metabolite nulling [Mader 2001]).

- *Refocusing pulses*. These pulses invert the transverse magnetization, performing a rotation of 180° of the magnetization vector and are the basis of the spin-echo (see section 2.3.2). It can be proven that a time-symmetric inversion pulse is a refocusing pulse, and vice-versa.

- *Saturation pulses*. Saturation pulses are used to eliminate the signal of a given region or frequency. These pulses are similar to excitation pulses but are designed to produce a considerable phase-shift in the selected spins and, consequently, no refocusing is applied. Saturation pulses are usually followed by spoiler gradients that further dephase the selected spins. Such pulses are used to eliminate undesirable sources of signal, such as signal originating from the regions adjacent to a selected slice or fat signals that may contaminate the signal. For more information, see section 2.5.5.

2.2.9 Adiabatic pulses

B_1 field inhomogeneities can change the flip angle produced by an RF-pulse and consequently impact the quality of the signal obtained with a pulse sequence. MRS is particularly sensitive to B_1-inhomogeneities, since even in perfect acquisition conditions the SNR of most of the metabolites of interest is already low. Adiabatic pulses are a subset of pulses that are particularly insensitive to B_1-inhomogeneities. In these pulses:

$$\vec{B_{eff}}(t) \cdot \frac{d\vec{M}(t)}{dt} = 0 \tag{2.28}$$

which means that the change of the magnetization vector ($\frac{d\vec{M}}{dt}$) is always perpendicular to the effective magnetic field, or in other words, that the magnetization vector always precesses

around $\vec{B_{eff}}$. In order to achieve this, the changes in the effective field need to be slow enough. This condition is the basis for the *adiabacity factor* Q [Baum 1985] that measures how adiabatic a given pulse is:

$$Q(t, \Delta\omega) = \frac{|\gamma \vec{B_{eff}}(t)|}{|\frac{d\vartheta}{dt}|} \tag{2.29}$$

where $d\vartheta/dt$ is the angular speed, with ϑ being the angle between $\vec{B_{eff}}$ and $\vec{e_z}$. A pulse is said to be adiabatic if $Q > 5$ ([Slotboom 1993]). In subsection 2.5.3 an MRS pulse sequence (SADLOVE/LASER) using adiabatic pulses is presented.

2.3 Magnetic Resonance Imaging (MRI)

2.3.1 Spatial-encoding and K-Space

In order to produce NMR-based images, i.e. MRI images, it is necessary to encode the spatial coordinates in the signal. This can be done using gradients to create phase and frequency spatial dependencies. In both types of encoding, a gradient in a given direction is switched on. In the case of *phase-encoding*, this is done before signal acquisition which will cause spins to accumulate different phase-shifts depending on their location. In contrast, *frequency-encoding* is performed by activating a "readout" gradient during acquisition, which will cause spins in different locations to be precessing at different frequencies while the signal is recorded. Considering a 2D slice and a spatially dependent property $\rho(x, y)$ that we aim to map, the corresponding phase and frequency encoded signal originating from a slice is given by:

$$S(t) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \rho(x, y) e^{i(\omega(x)t + \phi(y))} dx dy \tag{2.30}$$

where frequency-encoding was applied in the x-direction and phase-encoding in the y-direction. In order to acquire enough information to produce a NxM image, the same number of points needs to be collected. This can be achieved, for instance, with M experiments where, in each experiment, N points are acquired from the NMR signal. To guarantee that each experiment conveys different information, the phase gradient is changed between experiments. The change in the phase-encoding gradient is usually performed by changing the amplitude of the gradient, however, since the dephasing effect is given by $amplitude \times duration$ of the gradient, the duration could potentially be changed between experiments. This latter option becomes less practical, since it can potentially increase the time required for acquisition, as well as increase the minimum echo-time possible with the sequence (see the next section for more details 2.3.2). Using the described approach, the frequency and phase dependencies are given by:

$$\begin{aligned} \omega(x, n)t &= \gamma G_x x n \Delta T_x, \qquad -N/2 \leq n \leq N/2 \\ \phi(y, m) &= \gamma m \Delta G_y T_y y, \qquad -M/2 \leq m \leq M/2 \end{aligned} \tag{2.31}$$

where m and n represent the discrete coordinates of the collected data, ΔG_y is the amplitude increment of the phase-encoding gradient between two consecutive experiments, and ΔT_x is the time interval between two consecutive points of the discrete signal that is recorded in each experiment. To facilitate the notation, we introduce the k-space formalism, a coordinate system where the discrete coordinates k_x and k_y are given by:

$$\begin{aligned} k_x[n] &= \gamma G_x \Delta T_x n \\ k_y[m] &= \gamma \Delta G_y T_y m \end{aligned} \tag{2.32}$$

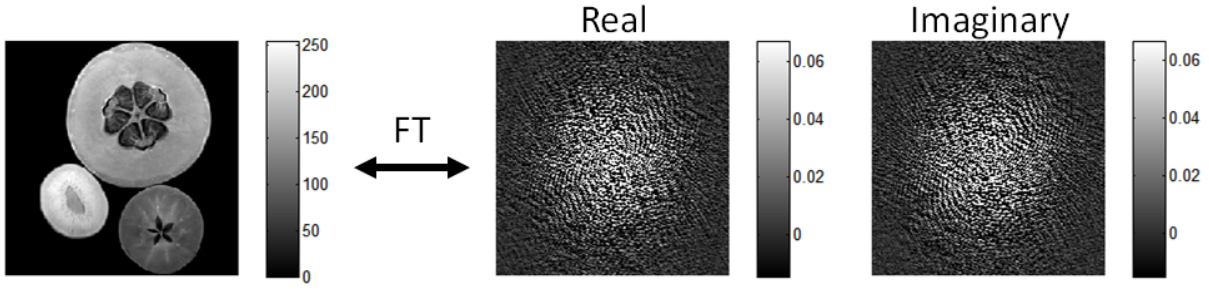

Figure 2.2: MRI image and corresponding k-space images (real and imaginary components). For visualization purposes the maximum value of the k-space color-scale was lowered, given that the amplitude of the center elements (low frequencies) are considerably higher than the rest of the k-space.

Therefore, the signal corresponding to a position $[n, m]$ of the NxM k-space image is given by:

$$S[n,m] = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \rho(x,y) e^{i(k_x[n]x + k_y[m]y)} dx dy \tag{2.33}$$

This means that for each k-space coordinate, the signal contains information from every spatial coordinate of the slice. Moreover, the value of $\rho(x, y)$ can be seen as the Fourier coefficient associated with the Fourier series element $e^{i(k_x[n]x+k_y[m]y)}$. Consequently, using a 2D Discrete-Fourier-Transform (DFT) it is possible to obtain the image in spatial coordinates corresponding to the acquired k-space image. Considering this, $\rho[x, y]$ can be calculated as:

$$\rho[x,y] = \frac{1}{N \cdot M} \sum_{n=-\frac{N}{2}}^{\frac{N}{2}} \sum_{m=-\frac{M}{2}}^{\frac{M}{2}} S[n,m] e^{-i(k_x[n]x + k_y[m]y)} \tag{2.34}$$

Figure 2.2 shows an example of a 2D MRI image and the corresponding real and imaginary k-space images.

Finally, some important relations between the dimensions and resolution of the k-space and its corresponding image should be highlighted. The size of the image (Field-Of-View (FOV)), and of the k-space (spatial frequency bandwidth (BW)) are defined as:

$$\begin{aligned} FOV_r &= N\Delta r \\ BW_r &= N\Delta k_r \end{aligned} \tag{2.35}$$

Where r represents a given spatially encoded direction and N is the corresponding number of encoding steps in the same direction. Considering now the relation between sampling frequency (ω_s) and maximum signal frequency (ω_M) given by Nyquist-Shannon sampling theorem:

$$2\omega_M < \omega_s \tag{2.36}$$

it is possible to relate BW and FOV:

$$BW_r = \frac{2\pi}{\Delta r} \Leftrightarrow \Delta r = \frac{2\pi}{BW_r} \Leftrightarrow \frac{FOV_r}{N} = \frac{2\pi}{BW_r} \tag{2.37}$$

which shows that FOV and BW are inversely related. This means, for instance, that a larger spatial coverage (larger FOV) using the same number of voxels can be obtained by reducing the spatial frequency bandwidth and the k-space spacing Δk_r. This can be seen as removing the the higher spatial frequency components of the k-space and using them to better cover the

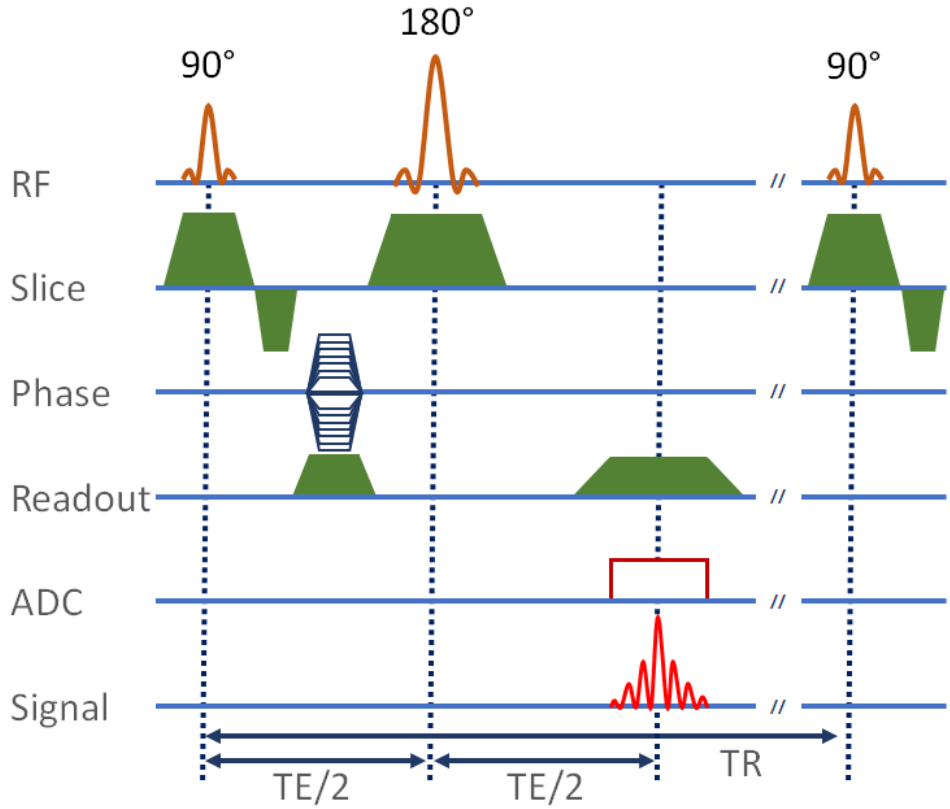

Figure 2.3: Schematic diagram of the spin-echo pulse sequence.

low-frequency region of the k-space, which becomes more important since the image is bigger and has a lower resolution. Similarly, the spatial resolution over the same FOV can be increased by covering a larger area of the k-space while maintaining the spacing Δk_r, i.e., adding more information concerning higher spatial frequencies.

2.3.2 Spin-echo sequence

The spin-echo (SE) sequence is one of the most fundamental pulse sequences in MRI and its basic elements are the basis for many of the pulses sequences currently used. The diagram of the sequence is presented in Figure 2.3 and in Figure 2.4 a schematic is shown showing the different steps that lead to the formation of the spin-echo.

The spin-echo sequence starts with a 90° excitation pulse that rotates the magnetization vector to the transverse plane (Figure 2.4, stage 1). To enable slice selection, a gradient is switched on at the same time as the 90° pulse. Directly after applying the pulse, the gradient is inverted and applied for half of the time to ensure that the phase-shift described previously in equation 2.27 is eliminated (Figure 2.4, stages 2 to 3). After this initial pulse, the effects related with T_2^* relaxation cause the different spins to dephase. The accumulated phase is given by:

$$\Delta\phi = \tau\Delta\omega \tag{2.38}$$

where $\Delta\omega$ is the difference between the frequency of each spin and the Larmor frequency, and τ is the time since the 90° excitation pulse. In order to rephase the spins, a refocusing pulse is applied that converts the accumulated phase from ϕ to $-\phi$ (Figure 2.4, stages 4 to 5).

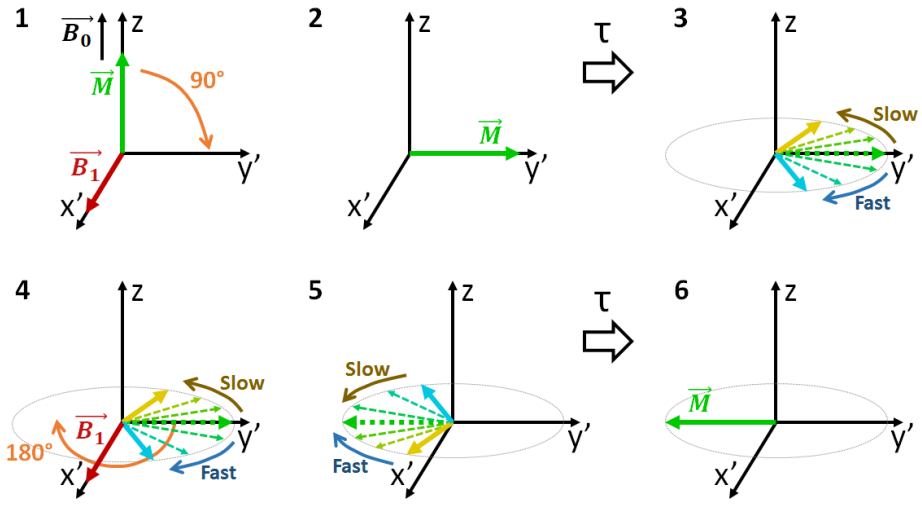

Figure 2.4: Different stages of the echo formation in a spin-echo sequence. For more details see description in the main text.

Consequently, after the same time interval as the one between the two RF-pulses, the spins refocus and a *spin-echo* is produced (Figure 2.4, stage 6). During the echo, the Analog to Digital Converter (ADC) is turned on and the signal is recorded. In a spin-echo sequence, the time between the magnetic center of the excitation pulse and the echo is called the echo-time (TE). The experiment can be repeated consecutively several times to allow for *spatial-encoding* or to improve SNR. The time between the initial excitation pulses of two consecutive experiments is called the repetition-time (TR). For spatial encoding, the phase-gradient is applied with a different magnitude in every experiment. Additionally, the readout gradient is switched on during acquisition. Given that the readout gradient will cause a phase shift, a preparatory inverted gradient is applied before the acquisition. In Figure 2.3 the preparatory gradient is applied before the 180° pulse and consequently has the same direction as the readout gradient. Due to the "unwinding" of the phase prior to acquisition, the echo will occur then the phase of the spins is fully refocused.

For the spin-echo sequence, the steady-state [1] relation between the signal amplitude and the tissue properties ρ (proton density), T_1 ,T_2, and sequence parameters TR and TE, is given by:

$$A(\rho, T_1, T_2, TR, TE) \propto \rho(1 - e^{-\frac{TR}{T_1}})e^{-\frac{TE}{T_2}} \tag{2.39}$$

Consequently, the weighting of the image in terms of the properties ρ, T_1 and T_2 depends on the values of TE and TR. Short TE and TR times are used to produce T_1-weighted images. In contrast, to create T_2-weighted images, long TE and TR times are required. Finally, proton-density (ρ) weighted images can be generated using long TR and short TE times.

2.3.3 Gradient-echo sequence

The gradient-sequence (Figure 2.5) is a sequence where the refocusing of the transverse magnetization is obtained using a gradient. In this sequence, the phase and readout gradients are switched on after the slice selection excitation pulse. These gradients lead to quick dephasing of the transverse magnetization that is refocused as result of the inversion of the readout gradient.

[1] Steady-state is the state reached after a few TRs, when the magnetization at the beginning of each repetition becomes constant.

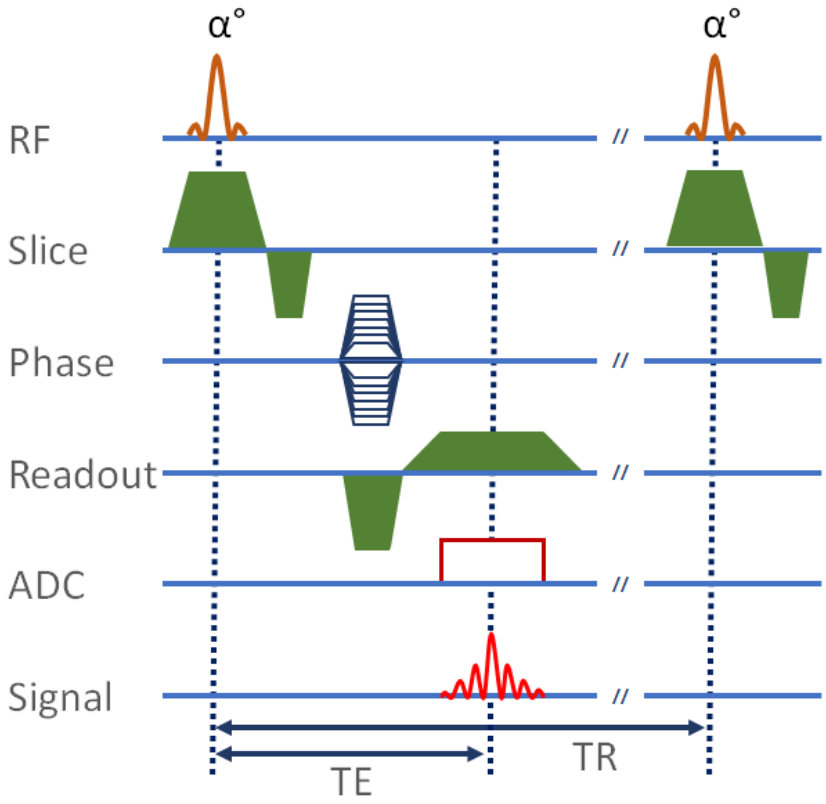

Figure 2.5: Schematic diagram of the gradient-echo pulse sequence.

In gradient-echo sequences, excitation pulses with a flip-angle smaller than 90° are often used. This leads to a lower SNR but allows a quicker recovery of the magnetization after each experiment, which allows for shorter TR values. For a given tissue having an average spin-lattice relaxation time of T_1 and a sequence with repetition time TR the angle that provides the best SNR is given by the Ernst relation:

$$cos(\alpha) = e^{-\frac{TR}{T_1}} \tag{2.40}$$

In the case of gradient-echo sequences, the flip-angle is often used to determine the contrast: large flip-angles are used to produce T_1-weighted images and short flip-angles T_2^*-weighted images. For large flip-angles, the higher the T_1 value the higher the saturation of the signal. In contrast, for small flip-angles the recovery of the longitudinal magnetization is quicker and signal saturation plays a less important role. In this case the main effect is related with T_2^*: signals with short T_2^* values will dephase quicker and, as explained, this sort of dephasing is not recovered by the field-gradients, leading to smaller signal amplitudes in those regions where T_2^* is shorter.

2.4 NMR Spectroscopy

2.4.1 Chemical-Shift

The external magnetic field is normally not directly experienced by the nuclei due to the electronic currents that it induces in the vicinity of each nucleus. These effects are proportional to the external field and, consequently, the local magnetic field is given by:

$$B_{local} = B_0(1 - \sigma) \tag{2.41}$$

where σ is the shielding constant, which has no units and is most of the times positive, i.e. the induced currents normally oppose $\vec{B_0}$, shielding the nucleus. For the same isotope, the shielding constant depends on the electronic structure in the vicinity of the nucleus, which changes depending on the molecular environment of the atom.

The change in the local magnetic field translates into in a shift of the Larmor or resonance frequency:

$$\omega_L = \gamma B_0(1 - \sigma) \tag{2.42}$$

Given that the change in frequency due to the shielding effects linearly scales with the external magnetic field B_0, the frequencies are often converted to *chemical shift* (δ) values which are given in the "ppm scale" and are independent of B_0. The conversion between a given resonance frequency ω and the corresponding chemical shift (δ) value is made using a reference signal and following the expression presented below:

$$\delta = \frac{\omega - \omega_{ref}}{\omega_{ref}} \times 10^6 + \delta_{ref} = \frac{\sigma_{ref} - \sigma}{1 - \sigma_{ref}} \times 10^6 + \delta_{ref} \tag{2.43}$$

Where δ_{ref}, ω_{ref} and σ_{ref} correspond to the chemical shift, resonance frequency and shielding constant of the reference signal. The multiplication by 10^6 (which explains the name "ppm = parts per million", not to be confused with ppm used to describe chemical concentrations), results from the differences in the resonance frequencies due to the shielding effects being usually extremely small compared to the Larmor frequency. Another important aspect of the ppm scale is that it is normally inverted, i.e. the numeric ppm values increase from right to

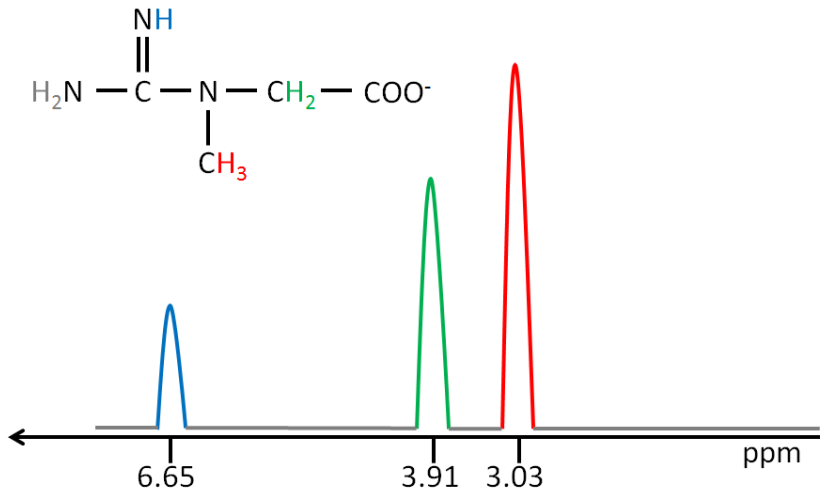

Figure 2.6: Image showing a schematic example of a proton NMR spectrum of Creatine, showing the different resonance peaks and the corresponding protons in the chemical structure (same color). Reference resonances from [Govindaraju 2000].

left. This is due to historical reasons, given that the first spectrometers searched for resonance signals by increasing the magnetic field strength, which is inversely related with the shielding constant.

One of the most commonly used reference signals is given by tetramethylsilane (TMS) which provides a strong signal due to its 12 equivalent protons. By convention, the ppm value of TMS has been defined as 0. Despite this, in *in vivo* applications water is most commonly used as reference signal ($\delta_{H_2O} = 4.7ppm$).

The dependency of the resonance frequency on the "neighboring" atoms, means that, in the same molecule, different nuclei (with non-zero spin) will have different frequencies depending on their location. Some of these nuclei might be equivalent and have the same resonance frequency. This can be the case for nuclei belonging to the same chemical group, as in $-CH_2$ and $-CH_3$ groups, where the protons of the same group are often chemically equivalent. The ensemble of resonances corresponding to a given molecule constitutes its NMR spectrum and given that, in most cases, is unique and can be regarded as the "chemical fingerprint" of the molecule. This feature is what enables NMR spectroscopy to be used for chemical composition analysis. Moreover, since for the same measurement conditions, the signal amplitude is proportional the number of nuclei (equation 2.17), NMR spectroscopy allows not only to detect the presence of a given compound but also to determine its quantity.

Figure 2.6 shows the proton NMR spectrum of Creatine, where the different resonance peaks can be seen. The area of the 3 peaks is proportional to the relative number of protons of each group. In this case, the relatively high electronegativity of the nitrogen, causes the proton of the $-NH$ group to be less "shielded" and consequently resonate at a higher frequency than the protons from the other groups. Another interesting aspect of the spectrum is that the protons of the $-NH_2$ group are not visible given that they exchange extremely rapidly with the solvent (water).

2.4.2 J-coupling

Besides the chemical shift, the magnetic moments of the different spins of the same molecule can influence each other and affect their resonance frequencies. In those cases, the spins are

said to be "coupled" and can not be described as two separate entities.

The effects of spin coupling can be described in terms of the energy levels of a group of nuclei and the corresponding transitions. For a system of two protons, A and X, with considerably different chemical shifts, there are 4 separate states. These states correspond to the different combinations of the individual states of the two spins, i.e.: $\alpha_A\alpha_X$, $\beta_A\alpha_X$, $\alpha_A\beta_X$, $\beta_A\beta_X$. Taking in consideration the energy of each spin:

$$E = -\gamma\hbar m_I(B_0 - \sigma) = -\hbar\omega m_I, \qquad m_I = -1/2, 1/2 \tag{2.44}$$

The energy of the system is given by the sum of the energies of both spins and the effect of spin-spin coupling:

$$E = -\hbar\omega_A m_{I,A} - \hbar\omega_X m_{I,X} + hJ m_{I,A} m_{I,X} \tag{2.45}$$

Where J is the scalar coupling constant in Hz, which reflects the strength of the interaction between the two spins and does not depend on the strength of the external magnetic field. The transitions associated with the different energy states can be given from expression 2.45, and considering that for every transition only one spin changes its state:

$$\begin{aligned} \Delta E_A &= \hbar\omega_A \pm \frac{1}{2}hJ \\ \Delta E_X &= \hbar\omega_X \pm \frac{1}{2}hJ \end{aligned} \tag{2.46}$$

Which means that the spectrum shows 4 peaks instead of 2. The 2 peaks corresponding to the transition of one nucleus are centered in the corresponding Larmor frequency and separated by J Herz.

For a AX_n, where X_n are equivalent nuclei, the signals corresponding to the resonances of the X spins only change in amplitude in proportion to the number of spins. In contrast, the A resonances are further split with increasing number of equivalent protons. Moreover, since the splitting always separates one peak in two new peaks centered in the original resonance and separated by J Hertz, some of the resonances may have more than one origin. As a consequence, the relative height of the resonances reflects the number of possible origins for each resonance. This means that the splitting pattern in function of n, leads to $1 + n$ peaks whose heights follow Pascal's triangle. The number of peaks originating from the splitting is called the multiplicity of the signal, and the naming convention depending on the number of peaks is: 1-singlet (s), 2-doublet (d), 3-triplet (t), 4-quadruplet (q), 5-quintuplet (quint), 6-sextuplet (sext), 7-septuplet (sept), >7-multiplet (m). Figure 2.7 shows the splitting patterns for different multiplicities.

Whenever a spin is coupled to more than one spin, the corresponding splitting pattern is a combination of the above-mentioned splitting patterns. For instance, if a spin A is coupled with two other spins with two different coupling constants, J_1 and J_2, a doublet of doublets (dd) is observed where the four lines correspond to the four different energy transitions given by:

$$\Delta E = \hbar\omega_A \pm \frac{1}{2}hJ_1 \pm \frac{1}{2}hJ_2 \tag{2.47}$$

Similarly, other combined patterns may be observed, such as doublet of doublet of doublets (ddd), triplet of doublets (td), etc..

The splitting due to spin-spin coupling described above hold whenever J is considerably smaller than the difference in resonance frequency between the two groups of coupled spins. This is called the "weak-coupling". In the remaining cases, when J is close to the difference in Larmor frequency of the coupled spins, the nuclei are said to be "strongly coupled" and the

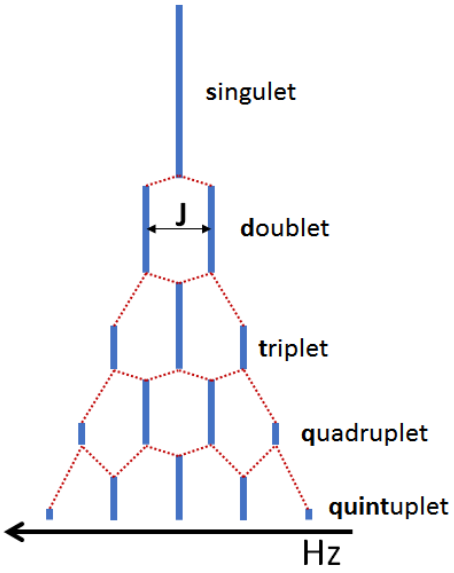

Figure 2.7: Splitting patterns originating from weak spin-coupling. The height of the peaks is proportional to its amplitude.

determination of the corresponding splitting patterns is considerably more complicated and often performed using computer simulations. Given that J does not depend on the magnetic field strength but the difference in resonance frequency does, a spin system can be either weakly or strongly coupled depending on the field strength.

2.4.3 J-coupling evolution (J-modulation)

The classical description used in the previous sections is adequate to describe the evolution of isolated spins in MRI-pulse sequences. However, the evolution of coupled spin systems can only be described using quantum mechanics. Despite a complete description of quantum mechanics being out of the scope of this thesis, a few important aspects necessary to understand the evolution of coupled spins are here introduced. For the interested reader, more information on the quantum mechanics description of NMR can be found in the source material of this section: [Rule 2006, Keeler 2004, Slotboom 1991].

In quantum mechanics, the state of a system is described probabilistically, using wave functions usually denoted as Ψ, that describe the probability of each state of the system. The probability density associated with a wave-function is given as $\Psi^*\Psi$, where Ψ^* is the complex conjugate of the wave-function. Measurable quantities are described using operators, that when applied to the wave-function, return the value of the observable multiplied by a wave-function. The evolution of the wave-function in time follows the principle of conservation of energy and is described by the Schrödinger equation:

$$i\hbar\frac{\partial}{\partial t}\Psi = \mathbf{H}\Psi \tag{2.48}$$

where $i = \sqrt{-1}$, $\hbar$ is the reduced Planck constant and $\mathbf{H}$ is the Hamiltonian operator which represents the total energy, depending the exact form of the Hamiltonian on the system being analyzed. In the specific case of NMR, the Hamiltonian is the sum of the Zeeman-energy, RF-pulse energy, influence of gradient fields, and J-coupling.

Given that the wave-function describes the probabilities of the different states, the observed quantities corresponding to a group of identical particles can be determined as the expectation values of the corresponding operators. For a given observable associated with an operator $\boldsymbol{A}$, the corresponding expected value A is given by:

$$A = \int_{-\infty}^{\infty} \Psi^* \boldsymbol{A} \Psi d\chi \tag{2.49}$$

where χ represents the coordinates of the coordinate-system in which the wave function is defined.

In NMR experiments, it is necessary to describe an ensemble of spins and their associated quantum states. In quantum mechanics, the description of an ensemble of quantum states is usually performed using the density matrix operator σ whose elements correspond to the average probability of all possible states of the ensemble. The expectation value of a given observable can be calculated from the density matrix as:

$$A = trace(\sigma \boldsymbol{A}) \tag{2.50}$$

The evolution in time of the density matrix can be derived from the Schrödinger equation and is given by the Liouville-von Neumann equations ([Canavagh 2007]):

$$\frac{d\sigma}{dt} = -i[\mathbf{H}, \sigma] \tag{2.51}$$

where $[\mathbf{A}, \mathbf{B}] = \mathbf{AB} - \mathbf{BA}$ is the commutator of the operators $\mathbf{A}$ and $\mathbf{B}$.

The density matrix for a system with a single spin (with $I = 1/2$) can be described using a linear combination of 4 operators: $\boldsymbol{U}$, $\boldsymbol{I_x}$, $\boldsymbol{I_y}$ and $\boldsymbol{I_z}$ where $\boldsymbol{U}$ is the identity operator and $\boldsymbol{I_x}$, $\boldsymbol{I_y}$ and $\boldsymbol{I_z}$ are the angular momentum operators in each spatial direction. Considering quadrature detection and the operator describing the transverse angular momentum $\boldsymbol{I_{xy}} = \boldsymbol{I_x} + i\boldsymbol{I_y}$, the FID can be obtained as the expectation value of $\boldsymbol{I_{xy}}$.

A system containing two coupled spins requires not only 4 operators but also the ones that account for the spin-spin coupling. The complete system is described by 16 independent operators: $\boldsymbol{U}$, $\boldsymbol{I_{1x}}$, $\boldsymbol{I_{1y}}$, $\boldsymbol{I_{1z}}$, $\boldsymbol{I_{2x}}$, $\boldsymbol{I_{2y}}$, $\boldsymbol{I_{2z}}$, $\boldsymbol{I_{1x}I_{2x}}$, $\boldsymbol{I_{1x}I_{2y}}$, $\boldsymbol{I_{1x}I_{2z}}$, $\boldsymbol{I_{1y}I_{2x}}$, $\boldsymbol{I_{1y}I_{2y}}$, $\boldsymbol{I_{1y}I_{2z}}$, $\boldsymbol{I_{1z}I_{2x}}$, $\boldsymbol{I_{1z}I_{2y}}$ and $\boldsymbol{I_{1z}I_{2z}}$. In this convention the subscript identifies the two spins and, as in the single spin system, $\boldsymbol{U}$ corresponds to the unitary operator. The terms $\boldsymbol{I_{1x}}$ and $\boldsymbol{I_{2x}}$ represent in-phase magnetization of spins 1 and 2 aligned along x. Similarly, $\boldsymbol{I_{1y}}$ and $\boldsymbol{I_{2y}}$ represent in-phase magnetization components for each spin along y. The operators $\boldsymbol{I_{1x}I_{2z}}$, $\boldsymbol{I_{1y}I_{2z}}$, $\boldsymbol{I_{1z}I_{2x}}$ and $\boldsymbol{I_{1z}I_{2y}}$ correspond to anti-phase magnetization. Only the in-phase operators are directly related with observable quantities. However, as will be shown, in coupled spins the anti-phase components are connected with the in-phase magnetization and consequently can be indirectly observed. The different operators and corresponding line-shapes are presented in Figure 2.8.

The Bloch equations for system of two strongly coupled spins can be derived from the Liouville-von Neumann equations, as described in [Slotboom 1994a], leading to the following

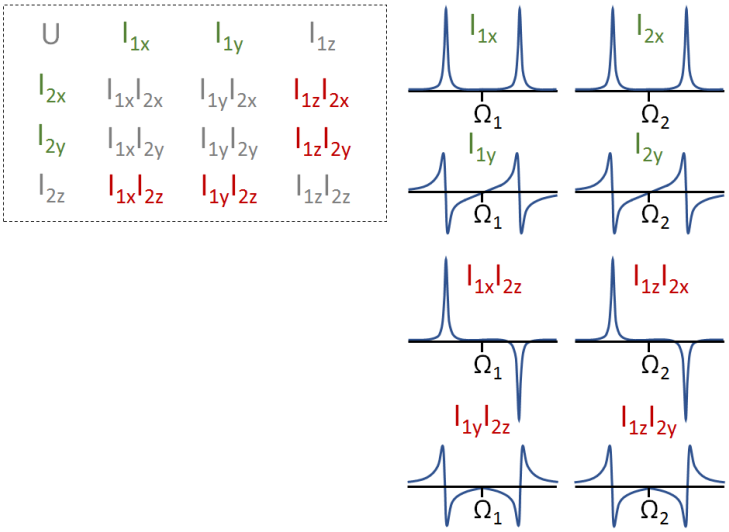

Figure 2.8: Different product operators of a two-spin coupled system. The in-phase operators were colored in green and the anti-phase operators in red. The operators in gray correspond to non-observable quantities. For all observable operators, the corresponding line-shape is displayed, where Ω_1 and Ω_2 correspond to the Larmor frequencies of each spin. Material used for image: [Keeler 2004, Rule 2006].

16 first order differential equations where relaxation was ignored:

$$
\begin{aligned}
\frac{dI_{1x}}{dt} &= -\omega_y I_{1z} + \Omega_1 I_{1y} - \frac{1}{2}JI_{1y}I_{2z} + \frac{1}{2}JI_{1z}I_{2y}^{\dagger} \\
\frac{dI_{2x}}{dt} &= -\omega_y I_{2z} + \Omega_2 I_{2y} - \frac{1}{2}JI_{1z}I_{2y} + \frac{1}{2}JI_{1y}I_{2z}^{\dagger} \\
\frac{dI_{1y}}{dt} &= \omega_x I_{1z} - \Omega_1 I_{1x} + \frac{1}{2}JI_{1x}I_{2z} - \frac{1}{2}JI_{1z}I_{2x}^{\dagger} \\
\frac{dI_{2y}}{dt} &= \omega_x I_{2z} - \Omega_2 I_{2x} + \frac{1}{2}JI_{1z}I_{2x} - \frac{1}{2}JI_{1x}I_{2z}^{\dagger} \\
\frac{dI_{1z}}{dt} &= -\omega_x I_{1y} + \omega_y I_{1x} + \frac{1}{2}JI_{1y}I_{2x}^{\dagger} - \frac{1}{2}JI_{1x}I_{2y}^{\dagger} \\
\frac{dI_{2z}}{dt} &= -\omega_x I_{2y} + \omega_y I_{2x} + \frac{1}{2}JI_{1x}I_{2y}^{\dagger} - \frac{1}{2}JI_{1y}I_{2x}^{\dagger} \\
\frac{dI_{1x}I_{1z}}{dt} &= \Omega_1 I_{1y}I_{2z} - \omega_x I_{1x}I_{2y} + \omega_y I_{1x}I_{2x} - \omega_y I_{1z}I_{2z} - \frac{1}{2}JI_{1y} + \frac{1}{2}JI_{2y}^{\dagger} \\
\frac{dI_{1y}I_{1z}}{dt} &= -\Omega_1 I_{1x}I_{2z} + \omega_x I_{1z}I_{2z} - \omega_x I_{1y}I_{2y} + \omega_y I_{1y}I_{2x} + \frac{1}{2}JI_{1x} - \frac{1}{2}JI_{2x}^{\dagger} \\
\frac{dI_{1z}I_{1y}}{dt} &= -\Omega_2 I_{1z}I_{2x} - \omega_x I_{1y}I_{2y} + \omega_x I_{1z}I_{2z} + \omega_y I_{1x}I_{2y} + \frac{1}{2}JI_{2x} - \frac{1}{2}JI_{1x}^{\dagger} \\
\frac{dI_{1z}I_{1x}}{dt} &= \Omega_2 I_{1z}I_{2y} - \omega_x I_{1y}I_{2x} + \omega_y I_{1x}I_{2x} - \omega_y I_{1z}I_{2z} - \frac{1}{2}JI_{2y} + \frac{1}{2}JI_{1y}^{\dagger} \\
\frac{dI_{1x}I_{1x}}{dt} &= \Omega_1 I_{1y}I_{2x} + \Omega_2 I_{1x}I_{2y} - \omega_y I_{1z}I_{2x} - \omega_y I_{1x}I_{2z} \\
\frac{dI_{1y}I_{1y}}{dt} &= -\Omega_1 I_{1x}I_{2y} - \Omega_2 I_{1y}I_{2x} + \omega_x I_{1z}I_{2y} + \omega_x I_{1y}I_{2z} \\
\frac{dI_{1z}I_{1z}}{dt} &= -\omega_x I_{1y}I_{2z} - \omega_x I_{1z}I_{2y} + \omega_y I_{1x}I_{2z} + \omega_y I_{1z}I_{2x} \\
\frac{dI_{1x}I_{1y}}{dt} &= \Omega_1 I_{1y}I_{2y} - \Omega_2 I_{1x}I_{2x} + \omega_x I_{1x}I_{2z} - \omega_y I_{1z}I_{2y} - \frac{1}{2}JI_{2z}^{\dagger} + \frac{1}{2}JI_{1z}^{\dagger} \\
\frac{dI_{1y}I_{1x}}{dt} &= -\Omega_1 I_{1x}I_{2x} + \Omega_2 I_{1y}I_{2y} + \omega_x I_{1z}I_{2x} - \omega_y I_{1y}I_{2z} - \frac{1}{2}JI_{1z}^{\dagger} + \frac{1}{2}JI_{2z}^{\dagger}
\end{aligned}
\tag{2.52}
$$

In these equations ω_x and ω_y are the frequencies of the x and y components of the amplitude-modulated pulse component, and Ω_1 and Ω_2 correspond to:

$$\begin{aligned} \Omega_1 &= \omega_1 + \omega(t) \\ \Omega_2 &= \omega_2 + \omega(t) \end{aligned} \tag{2.53}$$

where ω_1 and ω_2 are the Larmor frequencies of each spin, and $\omega(t)$ is the frequency of the frequency-modulated pulse component. The terms containing a † sign correspond to terms that are only relevant for strong-coupling and may be ignored for weakly-coupled spins. These equations show that the evolution of the magnetization of coupled spins is tightly connected to each other through the J-coupling terms. In the previously described Bloch equations derived using classical mechanics (equations 2.14) these terms are missing, showing that the classical description of NMR is unable to describe the evolution of coupled spin systems.

A manifestation of the effects described by these equations is seen, for instance, in the TE-dependent line-shapes of coupled spins. It can be shown that in a spin-echo the 180° pulse does not affect the evolution under J-coupling. Consequently, the conversion between in-phase and anti-phase magnetization leads to different line-shapes depending on the TE and the J-coupling constant. Pure in-phase magnetization is observed for $TE = m/J$, with $m = 1, 2, 3, \ldots$, being the signal negative for odd m numbers, and positive for even m numbers. In brain MRS, one of the most common examples of this is the doublet of lactate at 1.3ppm, which has a coupling constant of about 6.933Hz ([Govindaraju 2000]). This means that the lactate doublet is inverted for echo-times of about 144 ms, and shows a positive doublet for TE values around 288 ms. For very short echo times, since there is no time for J-evolution to occur, only positive in-phase magnetization is observed. Finally, it should be emphasized that at 1.5T most of the metabolites assessed *in vivo* contain spins that are strongly coupled.

2.5 *In vivo* Magnetic Resonance Spectroscopy (MRS)

In this section, we will introduce some of the most important pulse sequences used for clinical *in vivo* MRS.

2.5.1 PRESS sequence

The PRESS (Point RESolved Spectroscopy) sequence[Bottomley 1987] is one of the most widely available and commonly used sequences for *in vivo* MRS and is the sequence that was used to acquire the data explored in this thesis. The diagram of the PRESS sequence is shown in Figure 2.9. As it can be seen, the PRESS sequence shares some of the elements of the spin-echo sequence but includes some important adaptations. As in the SE-sequence, PRESS starts with a slice selective excitation pulse which is followed by a refocusing pulse, however, PRESS includes a second refocusing pulse. The working principle of PRESS is that only those spins that are excited and properly refocused in a given Volume of Interest (VOI) contribute to the localized signal. This is done by applying 3 orthogonal gradients, each one accompanying one of the three selective RF-pulses. During acquisition, no readout gradient is applied to guarantee that the observed differences in resonance frequency only originate from chemical-shift effects.

PRESS, as the other MRS sequences presented in this chapter, can be used to acquire a single spectrum originating from a single VOI or MRS voxel (Single Voxel Spectroscopy - SVS), or multiple spectra from multiple voxels contained in the spatially selected VOI (Magnetic Resonance Spectroscopic Imaging - MRSI). In MRSI, spatial encoding is performed by adding 2

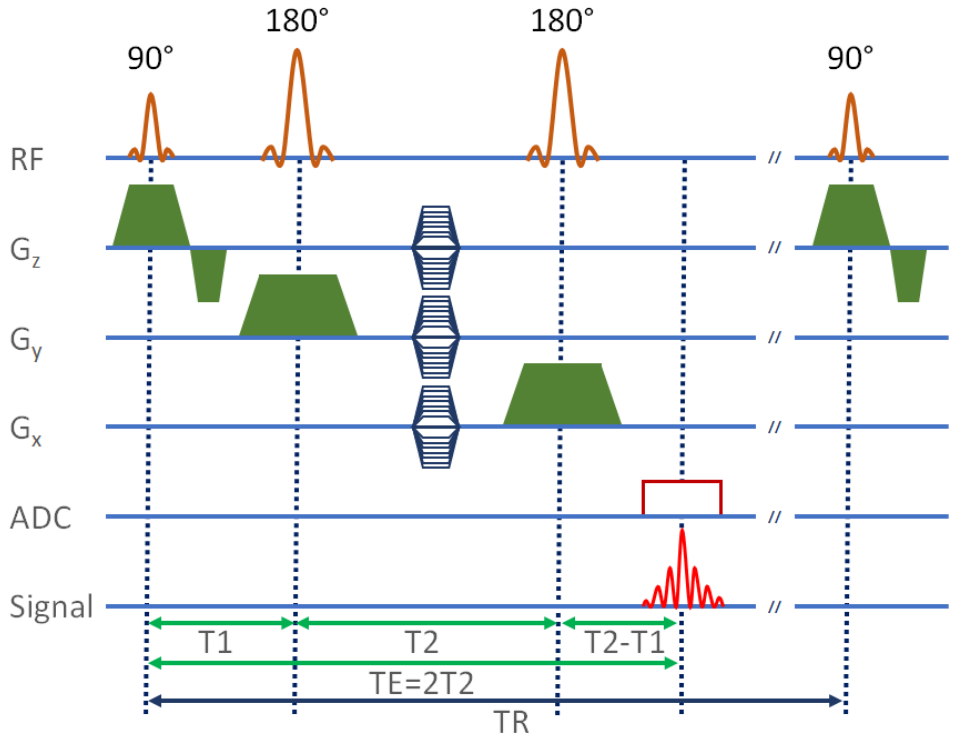

Figure 2.9: PRESS sequence diagram.

(or 3) orthogonal phase encoding gradients to the sequence. The use of no readout gradient means that, for every TR, instead of reading one line of k-space, only the information of a single k-space coordinate is collected. This difference makes PRESS considerably slower than conventional
acquisition speed translates into PRESS-MRSI being normally acquired with a considerably lower spatial resolution than structural MRI images.

Finally, the amplitude of the steady-state PRESS signal, not considering coupling effects, is identical to the spin-echo, i.e.:

$$A(\rho, T_1, T_2, TR, TE) \propto \rho(1 - e^{-\frac{TR}{T_1}})e^{-\frac{TE}{T_2}} \tag{2.54}$$

2.5.2 STEAM sequence

The STEAM (Stimulated Echo Acquisition Mode) sequence[Frahm 1987] is, together with PRESS, one of the most frequently used sequences for *in vivo* MRS. The main difference between STEAM and PRESS is that STEAM uses 3 90° selective excitation pulses and no refocusing pulses (see Figure 2.10). The first pulse rotates the magnetization vector into the xOy plane. After this pulse, the spins dephase due to T_2^* effects and a second selective pulse is applied which brings the magnetization into the z-direction. Assuming that before the second pulse the magnetization was equally distributed along the xOy plane, only about half of the magnetization is successfully brought along the z-direction. After the second pulse, the signal decays due to T_1 relaxation effects. Finally, the third pulse is applied which brings the magnetization back to the transversal plane. Given that the phase of the spins is inverted relatively to the moment immediately before the second pulse, a stimulated echo is created

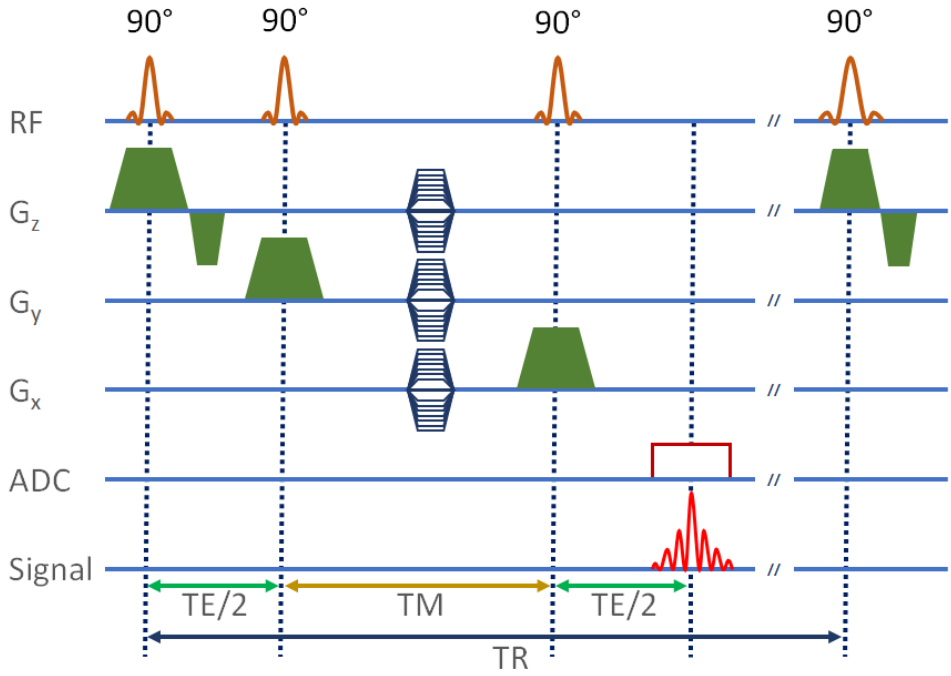

Figure 2.10: Schematic diagram of the STEAM pulse sequence.

after a period τ equivalent to the time between the first two pulses, i.e., $TE = 2\tau$. The period between the second and third pulses is called mixing time (TM). The signal amplitude of the stimulated echo is, without considering coupling effects, given by:

$$A(\rho, T_1, T_2, TR, TE) \propto \frac{1}{2} sin(\alpha_1) sin(\alpha_2) sin(\alpha_3) \rho (1 - e^{-\frac{TR}{T_1}}) e^{-\frac{TE}{T_2} - \frac{TM}{T_1}} \tag{2.55}$$

Since selective pulses have less SAR and a better selection properties, STEAM normally leads to smaller energy deposition and has better spatial selectivity than PRESS. However, these benefits come at the price of SNR reduction, which in theory is about half of that which can be achieved with PRESS for the same echo-time. The SNR and selection properties of PRESS and STEAM are further discussed in chapter 5, section 5.3. Another advantage of STEAM is that it allows for shorter TEs[Moonen 1989] since during the mixing time the signal does not decay with T_2, which means that less components need to be placed during the time equivalent to TE. Despite these advantages, the greater SNR of PRESS makes it superior for long TE acquisitions.

2.5.3 SADLOVE/LASER sequence

Non-adiabatic pulses, such as the ones used in PRESS and STEAM, require homogeneous B_1 fields to work properly. Therefore, given that at higher field strengths (>$3T$), homogeneous B_1 fields are more challenging to achieve[Ibrahim 2007], sequences using adiabatic pulses are desired. Moreover, when compared to conventional pure amplitude modulated pulses, adiabatic pulses have normally a considerably larger bandwidth at the same SAR. This allows to reduce the Chemical-Shift-Displacement-Artifact (CSDA), which increases linearly with the field strength and decreases with RF-pulse bandwidth. The CSDA is described in more detail in section 5.5.4.

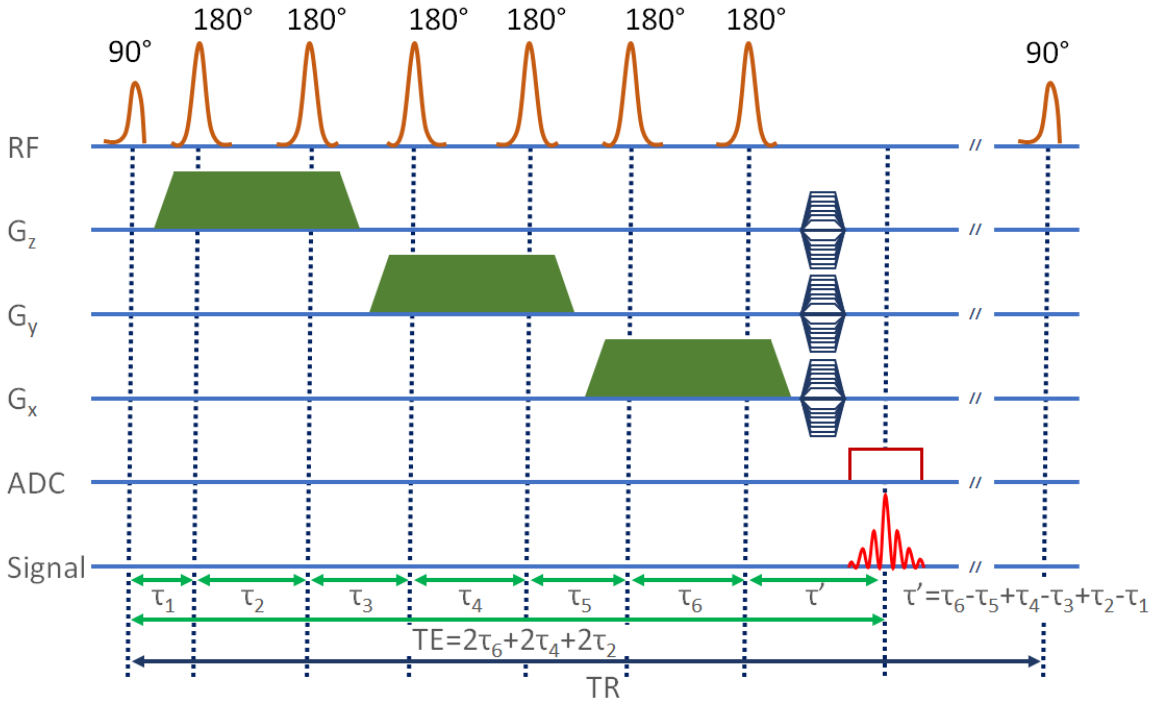

Figure 2.11: Schematic diagram of the SADLOVE (also known as LASER) pulse sequence.

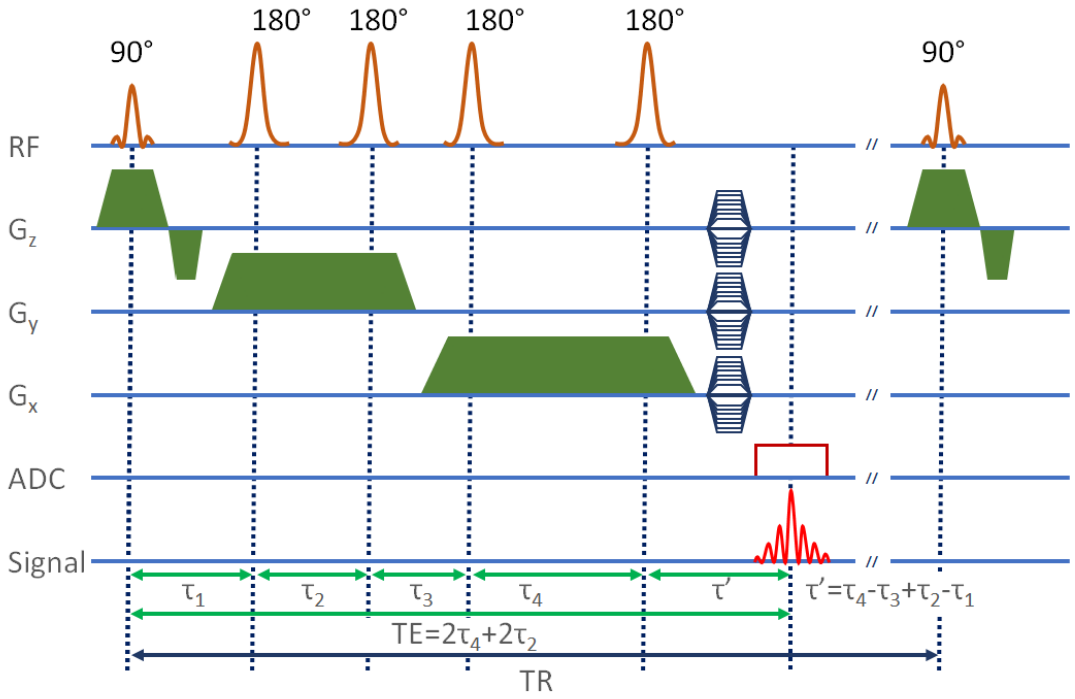

Figure 2.12: Variation of SADLOVE (also known as semi-LASER) sequence diagram.

In order to solve some of the issues with spectroscopy acquisition at high field strengths, a fully adiabatic localized MRS sequence was developed in 1991 by Slotboom et al. [Slotboom 1991]. The sequence, named Single shot ADiabatic LOcalized Volume Excitation (SADLOVE)[Slotboom 1991], was later on published with a different name (Localized by Adiabatic SElective Refocusing - LASER[Garwood 2001]) that is more commonly used. The sequence diagram of SADLOVE/LASER is presented in Figure 2.11. SADLOVE starts with an adiabatic non-selective half-passage excitation pulse that brings the magnetization to the transverse plane. Then, for each spatial direction a pair of identical 180° *sech/tanh* pulses is applied[Slotboom 1991]:

$$B_1(t) = \Omega_0 sech(\beta t)^{1+i\mu} \tag{2.56}$$

where Ω_0 is the pulse amplitude, β is the damping constant, μ is the selectivity constant and $i = \sqrt{-1}$. The need to use pairs of identical 180° pulses results from the non-linear offset frequency dependent phase variations produced by the first pulse, which are canceled by the second identical pulse[Zhu 2011]. Spatial selection is achieved by accompanying the 3 pulse pairs with 3 orthogonal selection gradients.

Besides the advantages of SADLOVE/LASER, i.e. insensitivity to B_1-inhomogeneities and low CSDA, the sequence does not allow for very short TE values, since all pulses and gradients need to fit in the interval corresponding to TE. To solve this, the first adiabatic excitation pulse can be substituted by a non-adiabatic slice selection pulse which enables removing one of the selective-refocusing pairs. This variation of SADLOVE/LASER is referred to as semi-LASER[Slotboom 1994b, Scheenen 2008] and the corresponding sequence diagram is presented in Figure 2.12.

2.5.4 SNR relation in MRS

In MRS, since the concentration of the metabolites of interest is usually very low, the pulse sequences are performed multiple times and relatively big selection volumes are normally used. If hardware and RF coil step considerations are ignored, the SNR relation for a given metabolite X can be approximated as described in ([Philips 2016]), by:

$$SNR \propto B_0 V[X]\sqrt{t_{total}} = B_0 V[X]\sqrt{N \cdot t_{acquisition}} \tag{2.57}$$

where t_{total} is the total measurement time, $t_{acquisition}$ is the acquisition time for each repetition, N is the number of repetitions, V is the volume and $[X]$ is the molar concentration of X. As shown in this equation, the SNR increases with $\sqrt{N}$ given that the signal scales linearly with the number of repetitions but the (stochastic) noise scales with the square root of the measurement time.

2.5.5 Water and Lipid Suppression

Water and lipids are considerably more abundant in the human body than almost every target metabolite that can be assessed *in vivo*. Consequently, water and lipid high signal amplitudes can obscure the proper assessment of other metabolite signals. In order to solve this, frequency selective pulses that only rotate the magnetization of spins whose chemical shift is within a narrow band can be used. Once these spins are in the transverse plane, crusher gradients are used to destroy the phase coherence of the selected spins. This technique is called CHESS

(CHEmical Shift Selective) and depending on the frequency band used for selection it can be used to eliminate the signals originating from water, lipids or any other resonances within a certain ppm range. In the brain, CHESS is often included as the first block of the spectroscopy sequence and is often applied in two different ways:

1. Water suppression - water suppression is normally performed in the whole FOV, i.e., no selection gradients are used when the CHESS pulses are applied.

2. Lipid suppression bands - lipid suppression is often performed to prevent the signals originating from the scalp to contaminate brain spectra. However, given that lipids (and lactate, whose major resonances are located in the same frequency range) contain relevant information for the assessment of several diseases, namely brain tumors, often the elimination of the lipid signals in the VOI is not of interest. Therefore, the chemical selective pulses are accompanied by selection gradients that allow the pulses to be applied only in certain regions.

2.6 MR Spectroscopy data explored in this Thesis

The work presented in this thesis explored a dataset of MRSI examinations collected from brain tumor patients. The acquisition was performed using the PRESS sequence (TE=135 ms and TE=30 ms, TR=1500 ms) combined with CHESS water suppression and using 1.5T Siemens scanners (models 'Aera' and 'Avanto', Siemens, Erlangen, Germany). The MRSI data had an original resolution of 12x12, with a voxel size of 13.33x13.33x15 mm, and was interpolated before DICOM storage to a resolution of 32x32 and a voxel size of 5x5x15 mm. Other details concerning the data used in each experiment/study are described in the corresponding chapters.

The measurements were performed in conformance to local and national ethical regulations, meaning that all patients gave their written informed consent that their data could be used for scientific purposes.

3 | MRS Data Processing & Machine Learning

After data acquisition, processing of MR spectra is required to enable the proper analysis of the acquired data. This chapter provides an overview of the most commonly used (pre) processing steps in MRS. The different methods presented here are available in the software tool jMRUI (`www.jmrui.eu`). From these, the majority is implemented in a novel jMRUI plugin named SpectrIm (Figure 3.1) that was developed during this thesis. Table 3.1 shows all different methods that were implemented in SpectrIm, as well as the methods from jMRUI that can be controlled from SpectrIm. Besides the methods shown in this table, a user may access all other jMRUI's methods by exporting the data from SpectrIm to jMRUI's main interface.

Most methods included in Table 3.1 are described in different sections of this chapter. However, the advanced methods that are considered to be the core of this Ph.D. thesis are presented in their own chapters, i.e.:

- Automatic quality control: chapters 6, 5 and 7.
- Brain Tumor Spectroscopic Imaging Segmentation (BraTSIS): chapter 10.
- Expected Distance to Tumor (EDT) maps: chapter 11.

All of the advanced methods developed during the Ph.D. programme are based on machine learning. Consequently, this chapter finishes with a section describing the main concepts of machine learning and the main machine learning algorithms explored in the next chapters of this thesis.

Table 3.1: Methods available in SpectrIm. The table shows which methods were implemented in SpectrIm and which methods from jMRUI can be can be controlled from SpectrIm through the batch processing interface. (X)-The two newest methods are not yet available in the distributed version of SpectrIm.

Method	Chapter/Section	Aim	SpectrIm	jMRUI
HLSVD	section 3.1.1	residual water removal		X
Automatic Quality Control	chapters 6, 5, 7	filter bad quality spectra	X	
Apodization	section 3.1.2	reduce noise	X	
Zero-filling	section 3.1.3	interpolate spectra	X	
Frequency-shift correction	section 3.1.4	align spectra	X	
Auto-phasing	section 3.1.5	adjust phase	X	
QUEST	section 3.1.6	quantification		X
BraTSIS	chapter 10	assist interpretation	(X)	
Expected Distance to Tumor	chapter 11	assist interpretation	(X)	

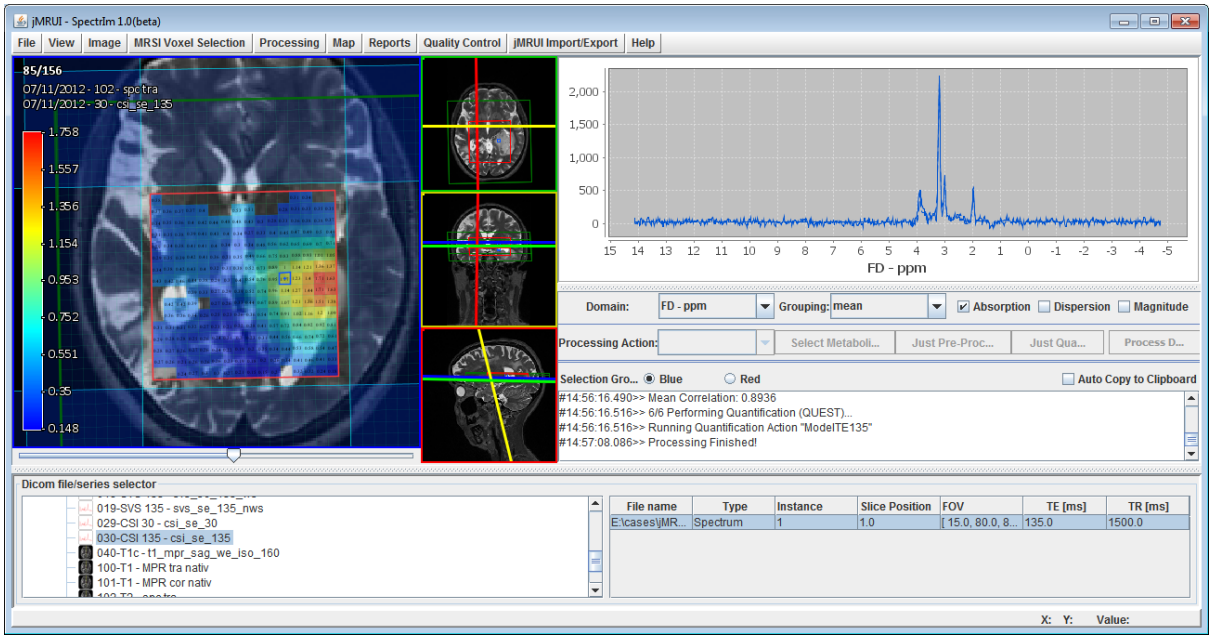

Figure 3.1: Main Graphical User Interface of SpectrIm. The figure shows a Cho/Cr map of a initial diagnostic acquisition of a GBM patient. MRSI data acquired at 1.5T, long TE (135 ms), PRESS.

3.1 Conventional MRS Post-Processing Methods

3.1.1 Residual water removal using Hankel Lanczos singular value decomposition (HLSVD)

Water signals have an amplitude several orders of magnitude greater than most of brain metabolites and, consequently, water-suppression techniques such as CHESS (section 2.5.5) are often required to enable the proper quantification of brain metabolites. However, in most situations water suppression is not complete, and a residual water signal is still left with an amplitude often superior to the amplitudes of the metabolites of interest. The residual water signal can influence the quality of quantification and, as it is discussed in chapter 6, the performance of automatic quality control. Thus, elimination of the residual water signal is an important post-processing step.

Residual water peak removal can be performed using the Hankel Lanczos singular value decomposition (HLSVD) method. HLSVD is a method used to decompose a time-domain signal $s(t)$, into K exponentially damped sinusoids:

$$s(t) \approx \sum_{k=1}^{K} A_k e^{-i(\omega_k t + \phi_k) - \alpha_k t} \tag{3.1}$$

where A_k, ω_k, ϕ_k and α_k are, respectively, the amplitude, frequency, phase and damping factor of the kth component.

For water removal, a frequency window for the exponential components corresponding solely to the residual water peak is set. Finally, once the different components with ω_k corresponding to the water resonance are estimated, residual water peak removal is performed by simply subtracting those components from the original signal.

The determination of the different components is made using the Lanczos algorithm [Pijnappel 1992, de Beer 1992], which is used to accelerate the Singular Value Decomposi-

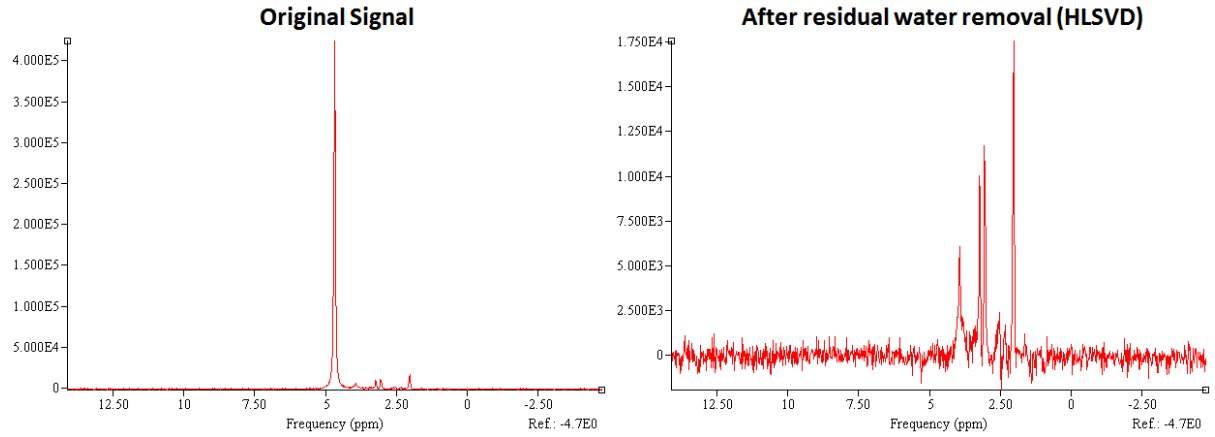

Figure 3.2: Proton NMR spectrum of the brain before and after residual water removal with HLSVD (jMRUI implementation).

tion (SVD) of the Hankel matrix form of the time-domain signal. Considering a time-domain signal $s = [s_1, ..., s_N]$ with length N, the corresponding $L \times M$ Hankel matrix H, with $L = N - M + 1$, is given by:

$$H = \begin{bmatrix} s_1 & s_2 & \dots & s_M \\ s_2 & \dots & \dots & \dots \\ \dots & \dots & \dots & s_{N-1} \\ s_L & \dots & s_{N-1} & s_N \end{bmatrix} \tag{3.2}$$

After SVD, the singular values correspond to the amplitudes A_k, and the singular vectors to the exponentially damped sinusoids.

The impact of the selection of the parameters N, M and K in HLSVD's performance for residual water removal is discussed in [Cabanes 2001], and the jMRUI implementation of the HLSVD algorithm is presented in [Beer 2015]. Figure 3.2 shows a brain spectrum before and after residual water peak removal.

3.1.2 Apodization

Apodization is a time-domain filtering technique used to improve the SNR of a spectrum. Signal filtering is performed by multiplying the time-domain signal with a weighting function, i.e.:

$$s'(t) = s(t) \cdot W(t, \theta) \tag{3.3}$$

The weighting function $W(t, \theta)$, where θ represents the parameters of the function, usually assigns higher weights to the points closer to the maximum of the time-domain signal, and lower weights to the points far from the maximum where the signal is dominated by noise. In SpectrIm two weighting functions, exponential and Gaussian, can be selected by the user:

$$W_{Gauss}(t, A_{Gauss}) = e^{-(t * A_{Gauss})^2} \tag{3.4}$$

$$W_{exp}(t, A_{exp}) = e^{-t * A_{exp}} \tag{3.5}$$

where A_{Gauss} and A_{exp} are, respectively, the Gaussian and exponential line broadening factors given in frequency units. The choice between these two filters depends on the type of line shape (Lorentzian, Gaussian or Voigt) considered in the quantification model.

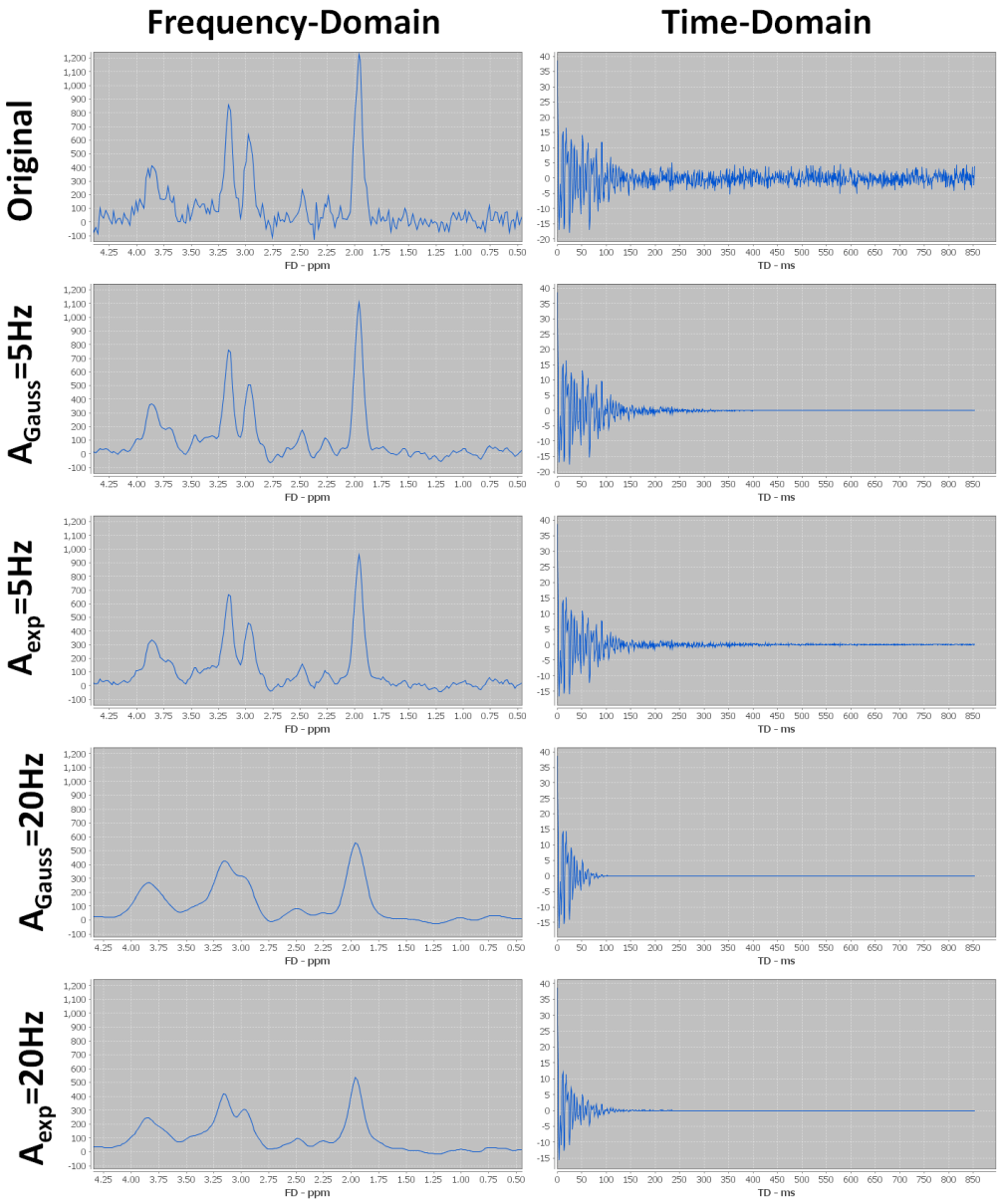

Figure 3.3: Frequency and time domain MRS signals (PRESS, 1.5T, TE=135ms) of the brain, before and after gaussian and exponential apodization with 5 and 20 Hz.

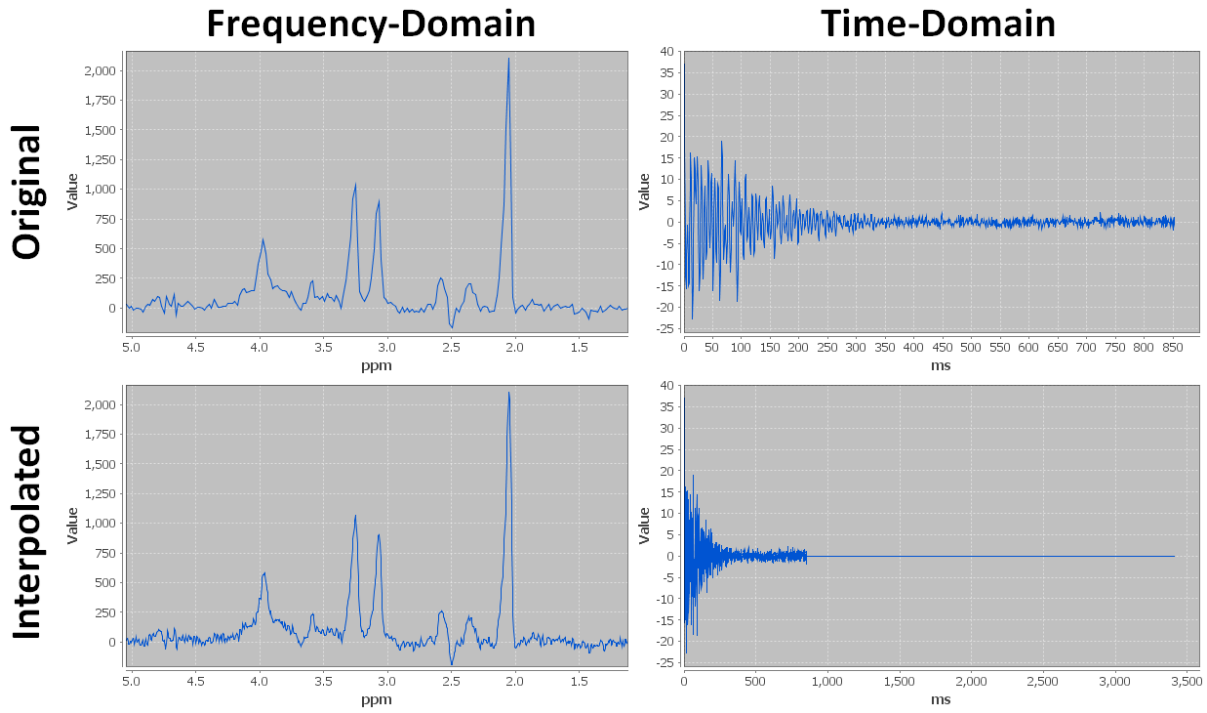

Figure 3.4: Zero-filling applied to increase the spectral resolution by a factor 4.

Figure 3.3 shows an example of spectra before and after apodization with a Gaussian and an exponential weighting functions. In these examples, the denoising effect is well visible, as well as the (undesirable) line-broadening that results from the apodization.

3.1.3 Zero-filling

Zero-filling, or Fourier interpolation, is an interpolation technique that explores the "scaling" properties of the Fourier transform.

In the previous chapter, the following relationship between bandwidth and spatial resolution was introduced in equation 2.35:

$$BW_r = \frac{2\pi}{\Delta r} \tag{3.6}$$

where r is a given spatial direction with a pixel spacing Δr, and BW_r is the bandwidth of the k-space image in the corresponding k-space direction. Similarly, for a time-domain signal:

$$T_{acq} = \frac{2\pi}{\Delta f} \tag{3.7}$$

where T_{acq} is the acquisition time and Δf is the frequency spacing of the spectrum. From this relation, it follows that an increase in the duration of the signal (T_{acq}) leads to an increase in the spectral resolution. Zero-filling is an interpolation technique where the duration of the signal is artificially increased by adding zeros to the end of the time-domain signal. After Fourier transformation of the zero-filled time-domain signal, an interpolated spectrum is obtained. Figure 3.4 shows the result of applying zero-filling to increase the spectral resolution of a spectrum by a factor 4.

Zero-filling can also be used to interpolate MR images by extending the k-space with zeros, as implied by equation 3.6.

3.1.4 Frequency-shift correction

B_0-drifts, as well as changes in temperature and *pH*, cause frequency shifts to the resonances of the NMR spectra. B_0-drifts affect all resonances equally and, consequently, if the reference frequency is adequately set, no changes in the *ppm* position of the peaks is observed. However, when B_0 changes and the synthesizer frequency is not adapted accordingly, the different metabolite peaks are recorded with a shifted frequency. Similarly, hardware problems that cause changes in the frequency of the synthesizer may lead to the same effect.

In contrast, frequency shifts due to temperature and *pH* changes affect different spins differently, causing resonance-specific ppm shifts. This effect can be used to determine the temperature from a proton NMR spectrum. In order to do so, the differences between the resonances of the main metabolites seen in the brain (water, Cho, NAA and Cr) can be used, as described in [Zhu 2008]. For *pH* measurements in the brain, however, either phosphorus spectra or proton spectra of exogenous compounds are used, given that the main metabolites visible in proton spectra do not provide enough information for measuring *pH* [Coman 2009].

Given that several methods rely on the exact ppm location of certain peaks to work properly (e.g. quantification and machine learning methods based on ppm-dependent features), the correction of the frequency shift is an important post-processing step. Most frequency-shift correction methods aim at correcting the *global shift* that affects all peaks of the same spectrum. Therefore, even after frequency shift correction some metabolite peaks might be displaced due to *pH* and temperature effects. As will be discussed further in this chapter, in the quantification section (section 3.1.6), quantification algorithms often allow the different peaks of the metabolite model to shift relatively to each other in order to accommodate metabolite-specific shifts.

Different approaches have been used to perform frequency-shift correction. In the INTERPRET project [Pérez-Ruiz 2010], for instance, an alignment algorithm was defined that used an hierarchical approach where a reference peak between choline, creatine and lipid (1.3), is selected based on the relative height. After selection of the reference peak, a Δppm shift corresponding to the difference between the location of the reference peak and its reference ppm value is applied to the spectrum. This method has the advantage of using fixed ppm references that are used to aligned all spectra, facilitating result comparison between different acquisitions. However, a very limited amount of information is used in the alignment process, namely only the position of the maximum corresponding to the selected reference peak, which may sacrifice the robustness of the method.

A different approach, named Spectral Registration, is proposed by Near et al. [Near 2015]. Spectral registration of a signal S is performed by least squares minimization of the difference between a reference signal R and the signal S after frequency and phase adjustment, i.e.:

$$\min_{\Delta\omega,\Delta\phi} \|R(t) - S(t)e^{i(\Delta\omega t+\Delta\phi)}\| \tag{3.8}$$

This method has the advantage of using more information than the previous given that the whole signal is used for alignment. However, it depends on the selection of an appropriate reference signal. In focal diseases, for instance, the spectra of the lesion may vary considerably relatively to a normal brain spectrum, possibly causing the alignment method to fail when the differences between the reference signal and the signal to be aligned are too big. In such situations, the Spectral Registration method can be adapted to allow for multiple reference signals. After performing the minimization with the different reference signals, the result with

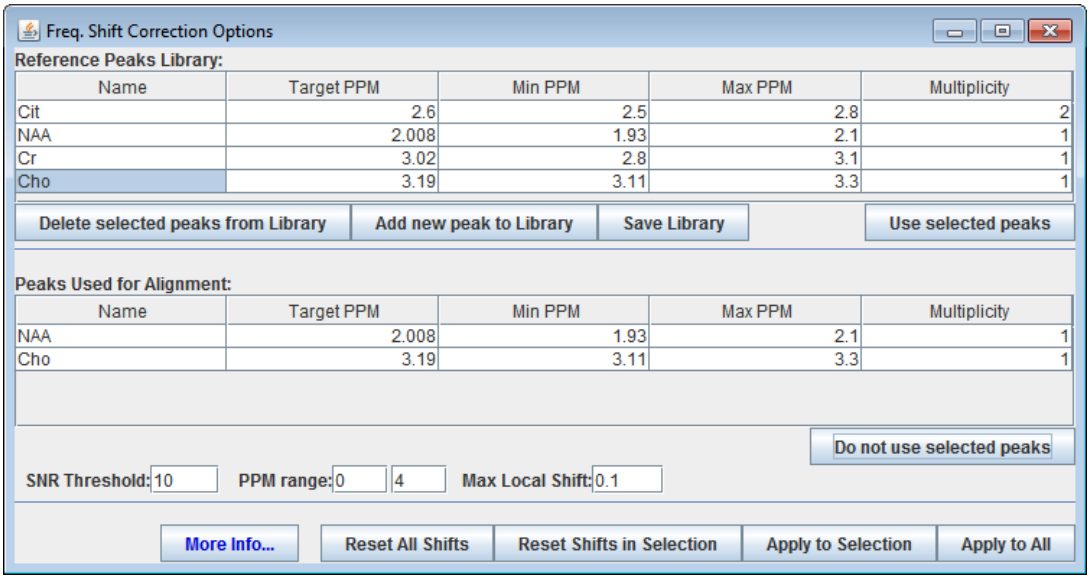

Figure 3.5: GUI of SpectrIm's frequency shift correction method.

the minimum residue is then selected, i.e.:

$$\min_{\Delta\omega,\Delta\phi,k} \|R_k(t) - S(t)e^{i(\Delta\omega t+\Delta\phi)}\| \tag{3.9}$$

where $R_k(t)$ is the k-th reference signal.

In SpectrIm, a novel (yet unpublished) method was implemented that combines some of the approaches used in the previous methods. The method was developed to allow robust alignment of MRS data and is composed of 3 steps, where steps 2 and 3 are MRSI specific:

1. Global alignment;
2. Local alignment;
3. Outlier correction.

In the *global alignment* step, the global shift ($\Delta\delta_{Global}$) is determined using the magnitude of the average spectra of all of the selected voxels. This allows for the global shift to be determined using a spectrum with higher SNR. Moreover, given that peaks with higher SNR are more reliable, the global shift is determined as the SNR-weighted average over the N detected peak-shifts $\Delta\delta_l$, where l is the index of the reference peak:

$$\Delta\delta_{Global} = \frac{\sum_{l=1}^{N} I(SNR_l)\cdot SNR_l\cdot\Delta\delta_l}{\sum_{l=1}^{N} I(SNR_l)\cdot SNR_l}, \quad I(SNR_l) = \begin{cases} 1, & SNR_l \geq T \\ 0, & SNR_l < T \end{cases} \tag{3.10}$$

In this expression, the indicator function $I(SNR)$ ensures that only peaks with an SNR above a defined threshold T are considered. The peak-shifts $\Delta\delta_l$ are calculated as the difference between the reference chemical shift of the peak and the observed chemical shift in the average spectrum. For peaks with odd multiplicity, the target chemical shift is the maximum value in a defined interval. In contrast, for peaks with even multiplicity, like lactate and citrate, the axis of symmetry is used instead and the reference δ value is provided accordingly. In SpectrIm, the information concerning the reference peaks can be stored in a *reference peak library* (Figure 3.5),

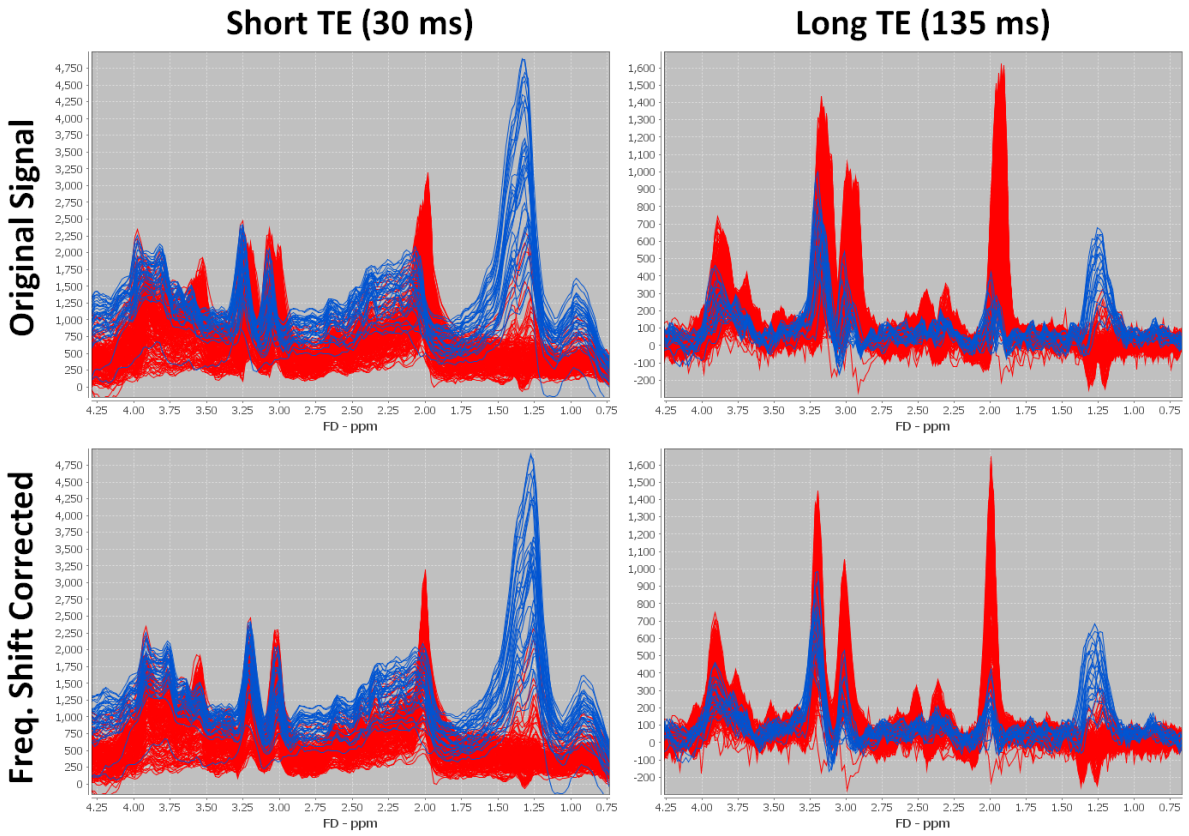

Figure 3.6: Short and long TE frequency domain signals before and after frequency shift correction using SpectrIm's algorithm (see description in main text). Different colors were used to discriminate solid tumor/necrosis spectra (blue) from the remaining spectra (red), showing that the method robustly corrects different types of spectra.

from which the user can select which peaks to use as reference for the alignment task. After determining $\Delta\delta_{Global}$, the shift is removed from the selected spectra.

The *local alignment* step aims at identifying the MRSI grid voxel-specific shifts, denoted as $\Delta\delta_{i,j,k}$, where (i,j,k) are the voxel coordinates. This is done by determining the local shift that maximizes the correlation between the magnitude spectrum of the voxel (i,j,k) and the magnitude of the average spectrum, i.e.:

$$\Delta\delta_{i,j,k} = \Delta\delta_{Global} + \underset{-K\leq\Delta\delta\leq K}{\arg\max}(corr(X_{i,j,k}[\delta-\Delta\delta], \bar{X}[\delta])) \tag{3.11}$$

where K is the maximum allowed local shift defined by the user.

Once all local shifts are determined, the *outlier correction* step is performed which ensures that there are no abrupt changes between the $\Delta\delta_{i,j,k}$ of neighboring voxels. Outlier correction is made using a Hampel Filter [Pearson 2016]:

$$\Delta\delta_{i,j,k} = \begin{cases} \Delta\delta_{i,j,k}, & |\Delta\delta_{i,j,k} - \tilde{\Delta\delta}_{i,j,k,R}| \leq C \cdot \underset{r\in[-R,R]}{MAD}(\Delta\delta_{i+r,j+r,k+r}) \\ \tilde{\Delta\delta}_{i,j,k,R}, & |\Delta\delta_{i,j,k} - \tilde{\Delta\delta}_{i,j,k,R}| > C \cdot \underset{r\in[-R,R]}{MAD}(\Delta\delta_{i+r,j+r,k+r}) \end{cases} \tag{3.12}$$

where C is a user defined constant and $\underset{r\in[-R,R]}{MAD}(\Delta\delta_{i+r,j+r,k+r})$ and $\tilde{\Delta\delta}_{i,j,k,R}$ are, respectively, the median absolute deviation and the median of the voxel-specific shifts in the neighborhood R around the voxel (i,j,k).

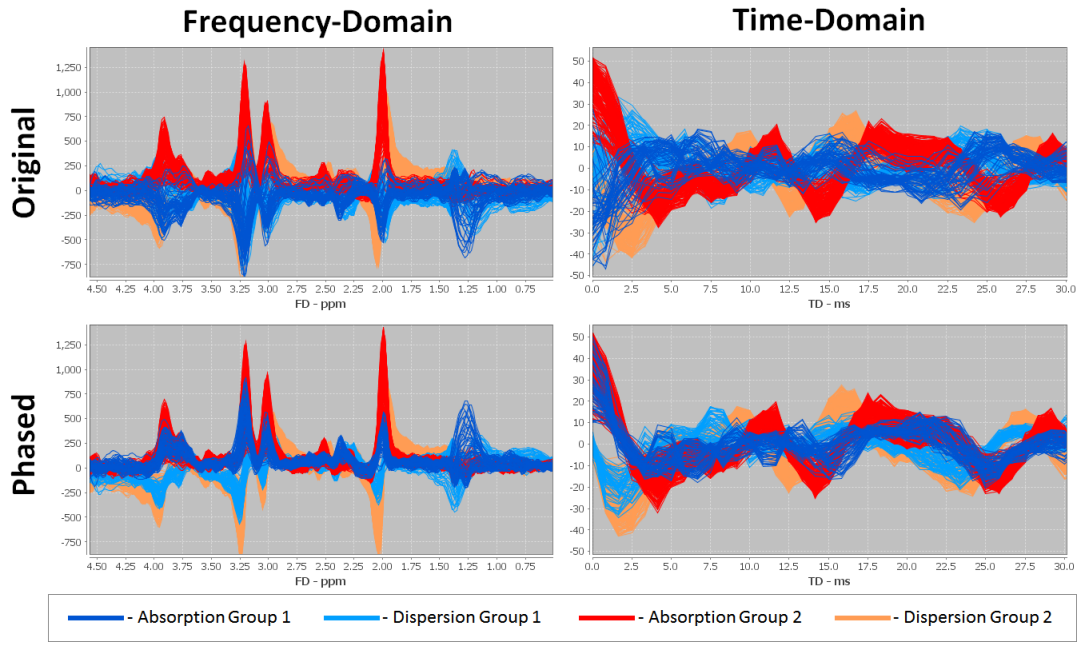

Figure 3.7: Selective auto-phasing performed using the method described by expression 3.15 in the chemical shift interval $[0,5]ppm$. The spectra of two different regions with different phase errors was colored differently to highlight their phase differences. All spectra belong to the same MRSI grid of a GBM patient. Acquisition details: PRESS, 1.5T, TE=135ms.

Figure 3.5 shows the GUI of SpectrIm's frequency shift method where the different parameters that can be set by the user are visible, and Figure 3.6 shows the spectra of an MRSI acquisition of a GBM patient, showing the robustness of the method to align different types of spectra.

3.1.5 Auto-phasing

As described in section 2.2.6, the NMR signal originates from an oscillation of the transverse components of the net magnetization, M_x and M_y. The oscillation is detected as two separate signals $s_x(t)$ and $s_y(t)$, which can be seen as the real and imaginary parts of a complex time-domain signal $s(t)$, with amplitude A, frequency ω and phase ϕ:

$$s(t) = s_x(t) + is_y(t) = A\cos(\omega t + \phi) + iA\sin(\omega t + \phi) = A \cdot e^{i(\omega t + \phi)} \tag{3.13}$$

where relaxation was not included and only zero-order phase was considered. Since the initial phase of the signal is often non zero, phase correction often needs to be performed. Zero-order phase correction of the whole signal (*global-auto-phasing*) can be performed using the first point of the time-domain signal:

$$s'(t) = e^{-i\phi}s(t), \quad \phi = \arccos\left(\frac{s_x(0)}{\sqrt{s_x(0)^2 + s_y(0)^2}}\right) \tag{3.14}$$

Global-auto-phasing of the signal can be performed whenever the different resonances of the signal have identical phases. However, in many cases this is not the case and it is interesting to selectively phase only some of them (*selective-auto-phasing*). This can be done by determining the phase that maximizes the sum of the absorption of the frequency-domain signal in a given chemical shift range $[\delta_{min}, \delta_{max}]$, i.e.:

$$\phi = \underset{-\pi<\phi\leq\pi}{\arg\max}\Big(\sum_{\delta=\delta_{min}}^{\delta_{max}} S_x(\delta)\cos(\phi) - S_y(\delta)\sin(\phi)\Big) \tag{3.15}$$

For this method to work properly, the limits of the chemical shift interval selected should not coincide with any peak. An example of the results of this method for the auto-phasing of all spectra of one MRSI grid are shown in Figure 3.7.

3.1.6 Quantification (QUEST)

Quantification aims at extracting metabolite quantities from MR spectra. Depending on the acquisition and quantification method used, quantification can result in absolute concentration values (absolute quantification), or be semi-quantitative, when the output is in the form of relative metabolite quantities. Absolute metabolite quantification, despite desirable, requires information that is not always available, such as: the tissue/compartment fractions for each voxel, the relaxation properties of the different metabolites in the each compartment, and the knowledge of the evolution of the spins for the sequence used ([Gaparovic 2016, Slotboom 1998]). Consequently, semi-quantitative methods are more frequently used in *in-vivo* clinical applications of MRS.

In this thesis, semi-quantitative quantification was performed using QUEST (QUantitation based on QUantum ESTimation, [Ratiney 2004, Ratiney 2005]). The QUEST algorithm performs quantification by modeling the acquired time-domain signal of dimension N, with sampling interval Δt, as a combination of M different known metabolite signals $\hat{x}^m$, as described by the following model:

$$\hat{x}[n] = e^{i\phi_0}\sum_{m=1}^{M} a_m \hat{x}^m[n] e^{(\Delta\alpha_m + i\Delta\omega_m)(n\Delta t + t_0) + i\Delta\phi_m}, \quad n = 1,\ldots,N \tag{3.16}$$

where $\hat{x}$ is the model estimate of the signal, $\hat{x}^m$ is the *m-th* metabolite signal of the pre-defined basis set, $i = \sqrt{-1}$, ϕ_0 is the global phase of the signal, t_0 is the dead time of the receiver, a_m is the amplitude of the *m-th* metabolite signal, and $\Delta\alpha_m$, $\Delta\omega_m$ and $\Delta\phi_m$ are the adjustments to the damping factor, angular frequency and phase shifts of the *m-th* metabolite signal. Therefore, QUEST receives as input the different signals $\hat{x}^m$, and estimates the relative amplitudes of the different metabolites, as well as the other described model parameters. The estimation is made by minimizing the squared euclidean distance between the estimated and the acquired signals using the Levenberg-Marquardt algorithm, i.e.:

$$\hat{\theta} = \min_{\theta} \|x - \hat{x}(\theta)\|^2 \tag{3.17}$$

where θ is the vector of the $4M + 2$ parameters estimated by QUEST:

$$\theta = [\phi_0, t_0, a_1, \ldots, a_M, \Delta\alpha_1, \ldots, \Delta\alpha_M, \Delta\omega_1, \ldots, \Delta\omega_M, \Delta\phi_1, \ldots, \Delta\phi_M] \tag{3.18}$$

Besides the parametric model of expression 3.16, QUEST is also able to estimate the macromolecular background signal. This is done under the assumption that the background signals decay rapidly and are mainly present in the first points of the time-domain signal. Consequently, in order to estimate the background signal, the first L points of the signal are truncated, and the previously described parametric model is fitted to the truncated signal. Then, the signal defined by the fitted parameters is extrapolated back to $t = 0$, which results in a signal

estimate $\hat{x}_{Met}$. The estimate of the background signal $\hat{b}$ is given by the subtraction between the original signal, x, and the estimate of the signal based on the truncated signal, $\hat{x}_{Met}$, i.e.:

$$\hat{b} = x - \hat{x}_{Met} \tag{3.19}$$

The estimate of the background, $\hat{b}$, can then be used in two different manners: *InBase* and *Subtract*. In the first approach, $\hat{b}$ is added to the basis-set as the M+1 metabolite, and QUEST is executed again to model the original acquired signal. In the second approach, the determined background signal is subtracted from the original signal and the resulting signal is then modeled using the described parametric model.

3.2 Machine Learning

Machine learning is, as described by Arthur Samuel in 1959, the *"field of study that gives computers the ability to learn without being explicitly programmed"*. A more precise definition of machine learning was given later, in 1997, by Tom M. Mitchel:

"A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P, if its performance at tasks in T, as measured by P, improves with experience E".

Machine learning has allowed to solve problems that, given their degree of complexity, couldn't be solved in the past using traditional programming, where the behavior of the program for every situation is explicitly defined. One of the most famous examples of such a problem is the development of bots, i.e. software agents that act as players, for the board game Go. Currently, the most successful Go bot is AlphaGo, the computer program that defeated the world champion of Go for the first time. As explained in the original paper [Silver 2016], the number of possible sequences of moves in Go is approximately 250^{150}, making it virtually impossible, using current computers, to compute the optimal move by exhaustive search. AlphaGo works by learning how to evaluate board positions and to select moves. This is done using reinforcement learning, a type of machine learning methods that is described below. The first versions of AlphaGo were initialized by training from human data and improved by self play. However, the most recent version of AlphaGo, Alpha Go zero [Silver 2017], has learned completely from scratch, starting with random play and learning by playing against itself.

Due to the increasing amount of available data and computing power, in part due to the advent of smart-phones, machine learning has become increasingly popular in the most recent years, as shown by the increasing google searches on the topic (Figure 3.8).

Machine learning methods can be distinguished in 5 main categories (`https://en.wikipedia.org/wiki/Machine_learning`):

1. **Supervised Learning** - Methods that learn from labeled data a function that, for a given input, predicts the output. Depending on whether the output is a continuous or a discrete variable, the learned model is referred to as, respectively, a regressor or a classifier.

2. **Semi-Supervised Learning** - Methods that aim at improving the performance of supervised learning using, besides the available labeled data, additional unlabeled examples.

3. **Active Learning** - Methods that aim at improving the learning efficiency of supervised methods by allowing the model to actively select the examples that carry the most information for training.

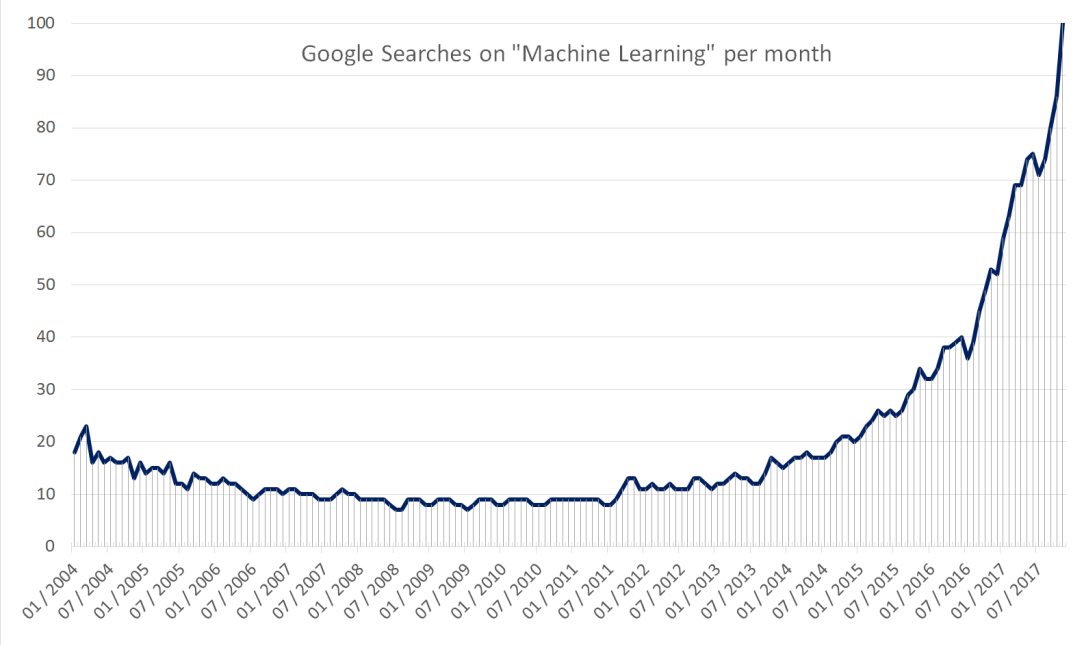

Figure 3.8: Popularity of the searches on "Machine Learning" in the past years. The y-axis shows the relative popularity index, which corresponds to the % of total number of searches in a given period of time on the topic "Machine Learning" rescaled to [0,100]. Source: Google Trends, `https://trends.google.com/trends/explore?date=all&q=%22Machine%20Learning%22`

4. **Reinforcement Learning** - Methods with a predefined goal that learn through its interaction with a dynamic environment.

5. **Unsupervised Learning** - Methods that learn the structure of unlabeled data. The most common example of unsupervised learning corresponds to clustering algorithms that identify clusters or groups in a given dataset.

In this thesis, different machine learning algorithms were used to develop novel MRS processing methods that aimed of tackling some of the difficulties associated with the clinical use of *in vivo* MRS. The main machine algorithms explored in this thesis, i.e. random forests, cNMF and X-means, are described in the next sections of this chapter. Furthermore, the machine-learning based methods for MRS data processing are described in the next chapters, more precisely:

- *Chapter 6* describes the use of a *supervised method* that is capable of learning to discriminate "good" quality from "bad" quality spectra based data labeled by two spectroscopists. The presented method is based on the *Random Forests classifier* [Breiman 2001, Liaw 2002]. Chapter 6 presents the results of the method for long TE data. The results for both short (30 ms) and long TE (135 ms) data are described in chapter 5.

- *Chapter 7* continues the work on automatic quality control and evaluates the use of an *active learning* approach to increase the efficiency of the labeling required to train the automatic quality control classifier.

- *Chapter 8* presents the results obtained with an *unsupervised method* named convex Non-Negative-Factorization (cNMF) ([Ding 2010, Ortega-Martorell 2012b]) that is used to learn a basis of spectra. The determined vectors are then used to perform fuzzy clustering

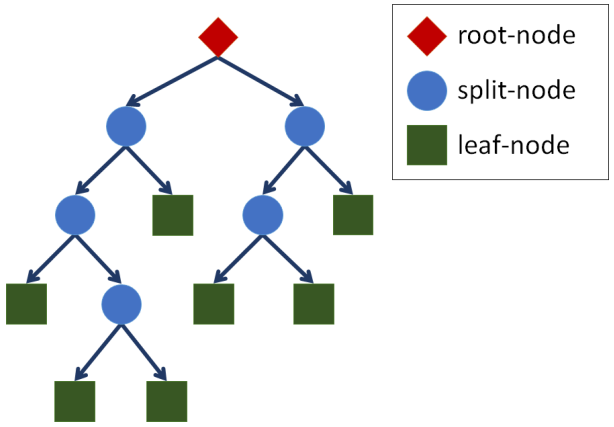

Figure 3.9: Schematic representation of a decision tree and its nodes.

of the spectra. The method was used to highlight differences between the spectra of GBM and metastasis patients.

- *Chapters 9 and 10* describe the use of an *(unsupervised) clustering method*, named X-means [Pelleg 2000] that is used to identify the main groups of spectra existent in MRSI examinations of GBM patients.

- *Chapter 11* describes the use of a supervised method, named Regression Forest [Breiman 2001, Liaw 2002], that is used to estimate the distance to the solid part of the tumor based on features extracted from individual spectra.

3.2.1 Random-Forests

Random forests (RF) [Breiman 2001] are supervised learning models that consist of an ensemble of randomly generated decision trees (Figure 3.9). The RF algorithm explores the concept of *bagging* (bootstrap aggregating, [Breiman 1994]), in which a prediction is made by averaging several versions of the same model, trained using different bootstrap[1] samples of the training set.

Bagging is used to improve the accuracy of a model by reducing the variance[2] of its predictions. Therefore, bagging is particularly interesting when applied to models with high variance and low bias, such as decision trees, given that the bagging reduces the prediction variance but not the bias [Hastie 2001]. The variance reduction achieved by bagging is limited by the correlation between the averaged models, i.e. the greater the correlation, the smaller the reduction in variance. Consequently, the random forests algorithm was developed as an improvement to bagging where the correlation between trees is decreased by injecting randomness in the creation of the decision trees. More precisely, in random forests, the input feature that is used in each node to split the data is chosen from a randomly selected subset of m features. This allows to reduce the correlation between the decision trees, and consequently increase the prediction accuracy in comparison with pure bagging approaches[Hastie 2001].

[1] A *bootstrap sample* is a sample drawn from a dataset using random sampling with replacement.

[2] Prediction *variance* and *bias* are two components of the prediction error. Variance refers to the square of the expected difference between each prediction and the expected value of the prediction. Bias, on the other hand, refers to the difference between the expected prediction and the true value being predicted.

The pseudo-code describing the random forests training process is presented in algorithm 1. As it is described in line 12 of the algorithm, after randomly selecting m features in each node, the feature that leads to the greatest reduction in "impurity" of the node is chosen. The lowest possible "impurity" after splitting is achieved when each child-node only contains examples of the same class. In contrast, the highest "impurity" level corresponds to a situation where the class distribution is the same in all child-nodes. Thus, the use of an "impurity" metric to drive the growth of the decision trees aims at generating leaf-nodes with members of only one class. For a given node i, the decrease in "impurity" resulting from splitting the node into two child-nodes is defined as:

$$\Delta F_i = F_{parent} - c_1 F_{child1} - c_2 F_{child2} \tag{3.20}$$

where F_{parent}, F_{child1} and F_{child2} are the values of a given "impurity" metric F measured in each node, and c_1 and c_2 are the weights of each child-node corresponding to the fraction of examples of the parent-node that were moved to each one of the child-nodes. Different "impurity" metrics for driving the selection of features in random forests are described in the literature [Hastie 2001, Criminisi 2013, Menze 2009]. For classification, the most common metrics are the *Gini impurity* and *entropy*, defined as:

$$Gini = \sum_{i=1}^{M} p_i(1 - p_i) \tag{3.21}$$

and

$$Entropy = -\sum_{i=1}^{M} p_i \log p_i \tag{3.22}$$

where M is the number of classes, and p_i is the fraction of elements of class i in a given node. For regression, the residual sum of squares (RSS) is used instead:

$$RSS = \sum_{i=1}^{N} (y_i - \bar{y})^2 \tag{3.23}$$

where y_i is the value of the output variable for the case i and $\bar{y}$ is the average value of the output variable in all examples of the node. In this case, given that the RSS scales with the number of examples in each node, the weights c_1 and c_2 of formula 3.20 are set to 1.

After training, the random forests model can be used to predict the output of a new example. This is done by taking the following steps:

1. In every tree, the test example is placed at the root-node and follows the path determined by the splitting rules defined during training ($\Theta_{i,j}$).

2. The information ($\Theta'_{i,j}$) collected from the training examples that, in each tree, reached the same leaf-nodes as the new test example is used to generate a prediction in the following manner:

    - In the case of *classification*, majority voting is usually performed, where each tree casts a vote. The vote corresponds to the majority class of the training examples that reached the same leaf-node as the test example.
    - In the case of *regression*, the prediction corresponds to the average value predicted by all trees. The prediction performed by each tree may correspond to the average value of the training examples that reached the same leaf-node as the test example. However, other approaches can be used, as described in [Criminisi 2013].

Algorithm 1 Random forests training pseudo-code.

for $i = 1$ **to** $nTrees$ **do**
  Draw a bootstrap sample S_i from the training set with the same size as the training set;
  Create a new tree T_i with a root-node $n_{i,0}$;
  Place the bootstrap sample S_i on node $n_{i,0}$;
  Create a set $LeafNodes = \{n_{i,0}\}$;
  Initialize node index counter: $k = 0$;
  while $LeafNodes \neq \emptyset$ **do**
    Select a node $n_{i,j}$ from set $LeafNodes$;
    if $NumberOfExamples(n_{i,j}) > minNodeSize$ **then**
      Create 2 new child-nodes $n_{i,k+1}$ and $n_{i,k+2}$ and add them to set $LeafNodes$;
      Select m features randomly;
      From the m features, select the feature that splits the data of node $n_{i,j}$ with the greatest reduction in "impurity"*;
      Split the examples using the selected feature;
      Move each part of the examples to the child-nodes $n_{i,k+1}$ and $n_{i,k+2}$;
      Save selected feature and decision rule parameters used for splitting as $\Theta_{i,j}$;
      Update node index counter: $k = k + 2$;
    else
      Save as $\Theta'_{i,j}$ the information of the training examples of node $n_{i,j}$ that is required for predicting the output of new test examples;
    end if
    Remove node $n_{i,j}$ from set $LeafNodes$;
  end while
end for

*- The different measures of node "impurity" are described in the main text.

As shown in algorithm 1, there are 3 main parameters of an RF model: (1.) the number of trees ($nTrees$), (2.) the minimum size of samples required for a node to be split ($minNodeSize$), and (3.) the number of randomly selected features per node (m). The selection of these parameters is normally made with the following considerations in mind:

- Increasing $nTrees$ leads to an asymptotic reduction in variance but a linear increase of the computational costs. Consequently, the number of trees is usually chosen at the point where the small gains in performance, obtained by further increasing the number of trees, do not justify an increase in computational costs. Depending on the problem, this number usually sits between 200 and 1000.

- The $minNodeSize$ controls the complexity of each tree and, therefore, smaller bias is usually associated with fully grown trees, i.e. $minNodeSize = 1$. However, when a considerable amount of noisy variables are present, fully grown trees might be highly unstable and may be prone to overfitting, namely when used for regression [Segal 2004]. Increasing $minNodeSize$ has a considerable impact on the computational costs, nevertheless, the prediction performance is usually the main factor driving the decision on $minNodeSize$. The selection of $minNodeSize$ is problem specific, however, in most cases, the optimization of this parameters leads to small gains in performance [Hastie 2001]. Consequently, this parameter is often not optimized. In the "R" implementation of random forests [Liaw 2002], the standard value for $minNodeSize$ is 1 for classification and 5 for regression.

- To understand the decision of the selection of m it is helpful to consider the two extremes first, i.e. $m = 1$ and $m = p$ where p is the total number of input features of the training set. In the first case ($m = 1$), the selection of the splitting feature in each node becomes purely random and consequently, the correlation between trees is minimal and the reduction in variance is maximal. However, given the lack of optimization of the features used in each node for splitting, the resulting decision trees will normally have a higher bias, which will govern the prediction error of the random forests model. In the second case ($m = p$), RF becomes essentially a pure bagging model. Consequently, the correlation between decision trees is maximal and the reduction in variance minimal. In this case, the bias is expected to be minimal, if overfitting is ignored, and the prediction error is governed by the variance. In summary, the selection of m corresponds to an optimization of a bias-variance trade-off and is problem specific. In the "R" implementation of random forests [Liaw 2002], the standard value for m is $\sqrt{p}$ for classification and $p/3$ for regression.

The simplicity and easy tunning of RF, combined with its elevated performance in various problems, has made random forests very popular [Hastie 2001]. Moreover, the RF provides some extra features that are extremely interesting in different problems: Out-Of-Bag (OOB) error estimates and variable importance assessment.

The OOB error estimates are made by measuring the prediction error of the examples left out of the bootstrap sample used for training each tree, i.e. using the OOB examples. In most situations, the OOB error is a good estimator of the prediction error of the classifier and come almost for "free", with a little extra amount of computation. However, in cases where the bootstrap sample is not completely independent of the OOB sample, the OOB error estimates may not generalize. In this thesis, for instance, OOB error estimates were not used given that the datasets were always composed of spectroscopic voxels from several MRSI examinations. Given that the sampling process of the random forests does not "know" the structure of the data, the voxels from each MRSI examination would be found, in most situations, both in the

bootstrap sample and the OOB sample, possibly leading to unrealistic performance estimations. The problem is aggravated by the fact that the different voxels are obtained after interpolation, meaning that the correlation between the voxels of the same examination is considerably high. Moreover, given that voxels of the same patient are likely to be correlated, even if originating from different examinations, the training and test sets used in the different projects presented in this thesis separated always the data into different datasets at the patient level, i.e. all the examples of the same patient were always placed in the same group (training, test or validation). Alternatively, this problem could've been solved by including information regarding the structure of the data on the sampling process that would sample the data at the patient-level, instead of the voxel-level. Such a strategy could also potentially increase the performance of the random forests models, given that it would reduce the correlation between different trees.

Regarding feature importance, the RF algorithm offers 2 types of measures to evaluate the relative importance of the input features: permutation-based and "impurity"-reduction-based. The first type corresponds to the average increase in error measured in each OOB sample after randomly permuting the values of a given feature between different Out-Of-Bag (OOB) examples. The greater the error increase, the greater the importance of the permuted feature for the predictions. The second type of measure assesses the contribution of each feature to split the data into leaf-nodes with members of only one class. This is done by summing, for each feature, the decrease in "impurity" (see equation 3.20) in all nodes where that feature was chosen for splitting. Thus, the features with the highest overall reduction in node "impurity" are deemed the most important. The feature importance measures described were used to analyzed different random forests models presented in this thesis (chapters 5, 6 and 11). The corresponding feature importance plots can be found in figures 5.7, 6.7 and 11.5.

3.2.2 cNMF

Non-negative-matrix-factorization (NMF) is a factorization method of the form [Ding 2010]:

$$X \approx FG^T \tag{3.24}$$

where:

- $X \in \mathbb{R}^{p \times n}$ is matrix representing the input data with n examples of size p.
- $F \in \mathbb{R}^{p \times k}$ is a matrix whose rows correspond to a set of k basis-vectors, named the *sources*.
- $G \in \mathbb{R}^{n \times k}$ is a *positive* matrix, named the mixing-matrix, whose elements correspond to the reconstruction factors used to describe X as a positive linear combination of the basis-vectors, or *sources*.

Convex-NMF (cNMF) is a variation of NMF, where the sources are determined as a convex combination[3] of the data points, i.e.:

$$F = XW \implies X \approx XWG^T \tag{3.25}$$

where $W \in \mathbb{R}^{n \times k}$ contains the coefficients that define the *sources* as a convex combination of input data vectors. Compared to conventional NMF, cNMF has several interesting properties, namely:

[3] a *convex* combination is a linear combination where all coefficients are positive and their sum is 1.

- The sources derived from cNMF are normally easier to interpret. In the case of the application of cNMF to factorize a set of MR spectra, this means that the sources are often "realistic" spectra, i.e. spectra that could potentially be observed. Conventional NMF, on the other hand, usually produces "unrealistic" sources that do not correspond to plausible MR spectra.
- In cNMF, the *sources* can be regarded as centroids, which allows interpreting cNMF as a clustering method such as K-means.
- The matrices W and G in cNMF are usually very sparse.
- In conventional NMF, the data matrix X is required to be positive, however, in cNMF, X is unconstrained.

A complete description of the cNMF algorithm is presented in [Ding 2010].

3.2.3 X-means

X-means [Pelleg 2000] is an extension of the clustering method K-means that defines a strategy to determine the number of clusters K. This is performed following the algorithm 2, where the minimum and maximum number of clusters, K_{min} and K_{max}, are user-defined model parameters.

Algorithm 2 X-means pseudo-code.

```
Initialize K = K_min;
while K ≤ K_max do
    Perform K-means until convergence;
    for each centroid c_i do
        BIC_i(k = 1) ←BIC score of the local data associated with c_i;
        Perform 2-means on the local data associated with c_i, leading to centroids c_{i,1} and c_{i,2};
        BIC_i(k = 2) ←BIC score of the local data after 2-means;
        if BIC_i(k = 2) < BIC_i(k = 1) then
            Substitute c_i with c_{i,1} and c_{i,2};
        end if
    end for
    Evaluate BIC-score of the clustering model considering all data points;
end while
Report model with highest overall BIC-score.
```

As it can be seen, the X-means algorithm is a hierarchical approach, where in each node a decision is made on whether to divide the local data into two new separate clusters using 2-means ($K = 2$) or not. The decision is made based on the Bayesian information criterion, defined as:

$$BIC = p\log(N) - 2\log\mathcal{L}(X|\theta) \tag{3.26}$$

where $\log\mathcal{L}(X|\theta)$ is the log-likelihood of the data X according to the K-means model parameters θ, p is the number of model parameters, and N is the number of data points of the data. BIC corresponds to measure of the quality of the fit of a model to the data (log-likelihood term), penalized by the model complexity ($p\log(N)$ term). The lower the BIC value, the better the

model. For local decisions on whether to subdivide a centroid, the data used to determine BIC corresponds solely to the data points associated with that specific centroid. In contrast, the overall BIC score calculated in step 12 of the X-means algorithm, considers all data points of the input dataset.

A more in-depth description of the X-means algorithm is found in the original paper by Dan Pelleg and Andrew Moore [Pelleg 2000].

4 | Brain Tumor Characterization using *in vivo* proton MRS

As we saw on chapter 2, Magnetic Resonance Spectroscopy (MRS) allows the retrieval of localized biochemical information from different parts of the human body in a non-invasive manner. The ability to detect metabolic changes in various tissues makes it a very appealing technique for diagnostics, treatment planning, and follow-up of many diseases. Currently, the main clinical applications of MRS are found in oncologic imaging[García-Figueiras 2016] of the brain[Preul 1996, Howe 2003, Dowling 2001, Majós 2004], prostate[Costello 1999, Mueller-Lisse 2007, Tayari 2017], and breast[Suppiah 2013, Cen 2014, Baltzer 2013], as well as in the evaluation of several metabolic diseases affecting the brain[Zimmerman 2011, Harting 2015, Razek 2014]. Other interesting applications include the assessment of neonatal hypoxic injury[Boichot 2006, Peden 1993] and the measurement of intramyocellular lipid content in skeletal muscle to assess insulin sensitivity in pre-diabetic patients[Boesch 1999, Boesch 2006].

This doctoral thesis is based on work performed in the Neuroradiology department of the University Hospital Bern. Consequently, the main clinical application explored in this thesis was brain cancer. Therefore, in the next two subsections, we first introduce the main brain metabolites measurable with MRS and, then, discuss what motivates the use of MRS for the clinical management of brain tumor patients.

4.1 Detectable Brain Metabolites in *in vivo* ^{1}H-MRS

Figure 4.1 shows a normal brain spectrum acquired using PRESS, where the resonances of the most important metabolites seen in healthy brain spectra can be seen. A brief description of the most important brain metabolites observed at 1.5T are described below.

4.1.1 Choline

Choline is an important constituent of cell membranes and myelin sheets, and is a precursor of the neurotransmitter acetylcholine [Govindaraju 2000, Mountford 2010]. In *in vivo* MRS, choline produces one of the most easily identifiable singlets in brain NMR spectra at 3.2 ppm. For simplicity, this resonance is frequently referred to just as choline, however, it originates from several metabolites[Govindaraju 2000], namely: free choline, glycerophosphorylcholine (GPC) and phosphorylcholine (PC). In the normal brain, the main contributors to the NMR visible choline correspond to PC (0.6 mM) and GPC (1 mM). Free choline is only found in small amounts in healthy brain tissue (<0.03 mmol/kg) but can be significantly elevated in tumors due to the increased membrane turn over of tumor cells.

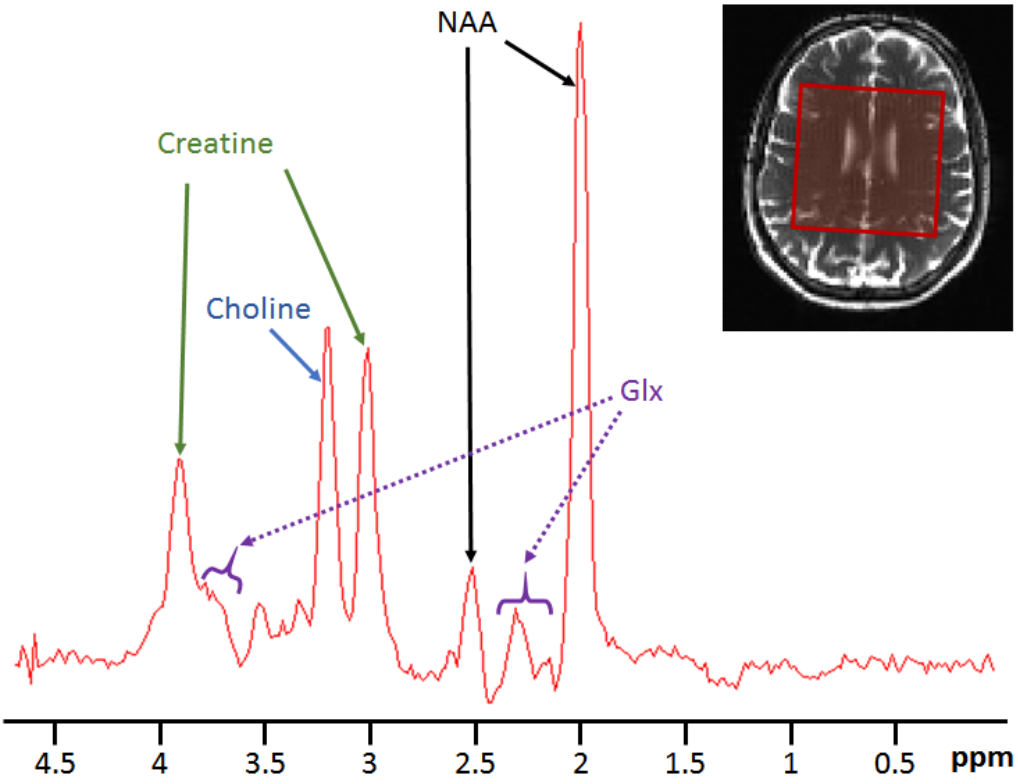

Figure 4.1: Normal brain spectrum acquired using PRESS (135 ms). The spectrum was collected from the region demarked with a red square in the reference image. The main visible peaks were assigned to the corresponding metabolites.

4.1.2 Creatine

Similarly to choline, the commonly referred creatine spectrum contains contributions from more than one metabolite, in this case, creatine and phosphocreatine. The main role of phosphocreatine is to act as an energy buffer[Greenhaff 2001], as it enables the process of converting adenosine diphosphate (ADP) into adenosine triphosphate (ATP), generating creatine in the process. Creatine is again converted into phosphocreatine in the sites of energy transduction, where ATP is consumed and converted into ADP. Creatine and phosphocreatine are characterized by two main resonances[Govindaraju 2000]: a singlet at 3.0 ppm from the methyl-protons and another singlet at 3.9 ppm from the methylene protons (see Figure 2.6). In the human brain, both metabolites are found in similar concentrations, with a total concentration of 6.4-9.7 mM in the gray matter and 5.2-5.7 mM in the white matter[Govindaraju 2000].

4.1.3 Glutamate and Glutamine

Glutamate and glutamine are two amino-acids found in the brain which, due to the overlap and similarity of their spectra, are frequently quantified together. The ensemble of glutamate and glutamine resonances is referred to as "Glx". Glutamate is the principal excitatory neurotransmitter in the vertebrate nervous system [Meldrum 2000] and is found in the brain with a concentration of about 12 mmol/kg, making it the most abundant amino acid found in the brain [Govindaraju 2000]. Glutamine's major role in the brain is a precursor to the excitatory neurotransmitters glutamate and aspartate, and the inhibitory neurotransmitter γ-amino butyric acid (GABA) [Albrecht 2010]. Both metabolites have distinct roles in tumor development. Glutamate is cytotoxic [Takano 2001], and its secretion induces an inflammatory response in the surrounding tissue, possibly caused by the neuronal loss resulting from glutamate itself. The inflammatory response is believed to contribute to tumor expansion[Takano 2001]. Glutamine, on the other hand, is a tumor nutrient and contributes to several essential metabolic tasks in tumor proliferation: it is involved in bioenergetics, protects the cells against oxidative stress, and complements glucose mechanism in the production of macromolecules [DeBerardinis 2010]. Glx resonances are found between 2.12 and 2.35 ppm, and between 3.74 and 3.75 ppm [Mountford 2010].

4.1.4 Glycine

Glycine is an inhibitory neurotransmitter and an antioxidant that is found throughout the normal brain with a concentration of about $1 mmol/kg_{ww}$ [Govindaraju 2000]. Proton NMR spectra of glycine shows a unique singlet at 3.55 ppm. The overlap between glycine and myo-inositol makes identification of glycine difficult at short TE, however, at long TE the myo-inositol resonances largely cancel each other enabling easier identification of the glycine peak. Glycine can be considerably increased in certain tumors, namely in a subset of gliomas [Choi 2011].

4.1.5 Lactate

Lactate is a product of anaerobic glycolysis and, consequently is a marker for reduced oxygenation [Mountford 2010]. In brain tumors, elevated lactate is often found in cystic compartments [García-Figueiras 2016]. Lactate produces a distinctive doublet at 1.33 ppm that, due to J-evolution, is inverted for echo-times around 144ms (see description of the J-evolution of the lactate doublet at the end of section 2.4.3).

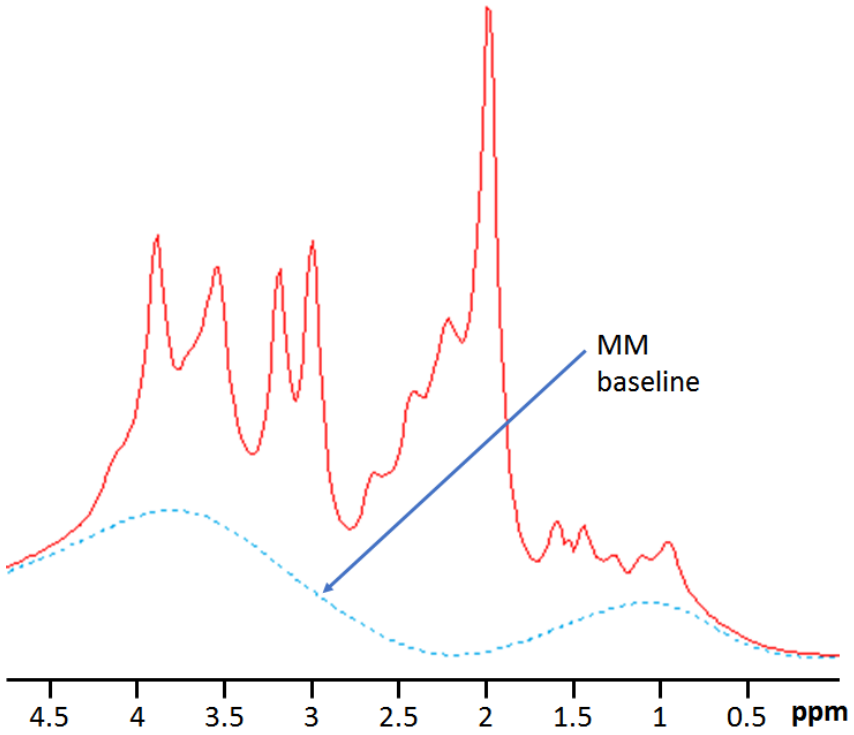

Figure 4.2: MR spectrum showing the macromolecular baseline in a healthy brain spectrum collected from the pons and a part of the surrounding tissues. The spectrum was acquired at 1.5T with TE=30ms and using PRESS. The MM baseline was calculated using QUEST [Ratiney 2005].

4.1.6 Macromolecular baseline

In *in vivo* MRS, the general classification of macromolecules (MM) is given to all molecules with high molecular-weight that, due to their short T_2 relaxation times, are difficult to resolve separately. The combined signals of all macromolecules seen in short TE spectra, forms the macromolecular baseline (see Figure 4.2). Macromolecule resonances can be separated in two main types [Cudalbu 2012]: (**1.**) resonances arising from the nuclei of atoms with extremely reduced mobility such as the ones forming the backbones of MM, and (2.) resonances from nuclei belonging to outer regions of macromolecules or to smaller MM that have a higher mobility than the nuclei from the previous group. In conventional MRS sequences, only the resonances corresponding to the second group are detectable, given that the signals from the first group decay extremely fast.

4.1.7 Mobile Lipids

The lipid signals observed at 1.3 ppm and 0.9 ppm correspond, respectively, to the $-(CH_2)$ and $-(CH_3)$ moieties of long chain fatty acids. Consequently, the ratio between the 1.3 and 0.9 signals is correlated with the length of the fatty acid chain ([Evaristi 2016]). In the case of unsaturated lipids, the $-CH=CH-$ produces a resonance at 5.4 ppm [Mountford 2010]. Similarly to the other macromolecules, the visibility of lipid signals in NMR is highly dependent on their mobility. Pospholipids contained in cell membranes, for instance, are highly immobilized and, consequently, have extremely short T_2 which renders them invisible in conventional MR spectra [Atlas 2009]. Lipids from the scalp, on the other hand, are more mobile and frequently contaminate brain spectra. Apart from lipid contamination from the scalp, the presence of

lipid signals in *in vivo* NMR spectra of the brain is often a marker of disease itself, and of tumor necrosis [García-Figueiras 2016].

4.1.8 Myo-inositol

Myo-inositol is an astrocyte marker and an osmolyte usually found in the brain with a concentration between 4 and 8 $mmol/kg_{ww}$ [Kreis 1997]. The spectrum of this metabolite is better detected at short TE and consists of 4 groups of resonances [Govindaraju 2000]: (**1.**) a doublet of doublet around 3.52 ppm, (**2.**) a triplet at 3.61 ppm, (**3.**) a triplet at 3.27 ppm that overlaps the choline peak, and (**4.**) a triplet at 4.05 ppm that is often affected by water suppression. Myo-inositol has been associated with gliosis, though the relation between high myo-inositol and gliosis is not clear [Hattingen 2008]. In gliomas, myo-inositol provides relevant information to predict tumor grade since it is often found elevated in low grade gliomas and decreased in high-grade gliomas [Metwally 2014, Castillo 2000].

4.1.9 N-Acetyl-Aspartate (NAA)

N-Acetyl-Aspartate (NAA) is an amino acid found in the healthy brain at high concentration levels (7.9-16.6 $mmol/kg_{ww}$ [Kreis 1997]). The acetyl moiety of NAA produces a singlet around 2.0 ppm which is one of the most prominent peaks in normal brain spectra (see Figure 4.1). Besides being relatively less visible than the acetyl singlet, the aspartate moiety leads to a few other resonances: two doublet of doublets at 2.67 and 2.48 ppm, another doublet of doublets at 4.38 ppm, and a doublet at 7.82 ppm from the amide group. NAA is considered a neuronal marker and low NAA levels are associated not only with neuronal loss (irreversible) but also with neuronal dysfunction that may be reversible in some situations [Govindaraju 2000].

4.1.10 Other detectable metabolites

In addition to the referred metabolites, other important substances can be detected in *in vivo* proton ^{1}H-MRS of the brain, such as:

- **Alanine** - a nonessential amino acid that is found increased, for instance, in meningiomas [Govindaraju 2000].
- **Gamma-Aminobutyric acid (GABA)** - the main inhibitory neurotransmitter of the central nervous system. [Puts 2012]
- **Glucose** - a simple sugar that is the main source of energy in humans. [Govindaraju 2000]
- **2-HG** - an oncometabolite seen in primary brain tumors that carry a mutation in the enzyme cytosolic isocitrate dehydrogenase (IDH) gene. IDH mutated gliomas are associated with longer overall survival. [Dang 2009, Choi 2012, Choi 2013, Andronesi 2012, Andronesi 2013]

Despite their importance, these four metabolites were not analyzed in the work presented in this thesis.

4.2 MRS in Brain Tumours

The ability to provide localized chemical information non-invasively makes MRS a very interesting tool for the metabolic assessment of brain lesions. As shown in the previous subsection describing the most relevant metabolites seen in *in vivo* brain MRS, many of the visible metabolites provide important information for the characterization of brain tumors. The use of MRS for the characterization of brain tumors can be divided into the following categories [Hollingworth 2006]:

1. differential diagnosis between tumor and non-cancerous lesion;
2. differential diagnosis between metastasis and high-grade tumor;
3. differential diagnosis between low-grade and high-grade tumor;
4. discrimination between tumor-progression and radiation necrosis;
5. delineation of tumor borders;

4.2.1 Tumor vs non-cancerous lesion

In some cases, conventional MRI imaging alone does not provide conclusive information to determine if a suspect lesion is cancerous or not. However, such distinction is extremely important given that certain benign non-cancerous lesions may not require a brain biopsy to be performed. Therefore, the correct differentiation between cancerous and non-cancerous lesions may eliminate unnecessary brain surgeries and all the associated risks [Horská 2010].

Figure 4.3 shows 4 different cases, 2 from brain tumors and 2 from lesions that can mimic brain tumors in MRI. As it can be seen in this particular example, the MR spectra of the 4 different cases are substantially different. For the AVM case, a tumor can be ruled out given that the lesion shows no brain metabolites and the surrounding tissue shows normal brain spectra. In the case of the bacterial infection, the presence of acetate, alanine, and amino acids at 0.9 and 3.75, which are potential markers of pyogenic abscesses [Lai 2005], enables the exclusion of a brain tumor. Regarding the brain tumor cases, the elevated metabolite ratios choline/creatine and choline/NAA are markers of malignant processes [Hollingworth 2006] and point towards a cancerous lesion. Despite this, diseases that lead to spectroscopic patterns with similar features as the ones observed in brain tumors, such as acute demyelinating lesions, may complicate the differentiation between malignant and non-malignant diseases using MRS alone [Horská 2010].

The performance of MRS to differentiate between neoplastic and nonneoplastic brain lesions was assessed in several studies. A study published in 2007 [Hourani 2008]. In that study, the authors showed that in a total of 69 brain lesions, including 36 brain tumors, 10 benign lesions, 4 strokes, 10 demyelination lesions and 9 stable lesions not confirmed on pathologic examination, it was possible to distinguish cancerous lesions from non-cancerous ones with an accuracy of 84.2%. Another study from 2002 [Möller-Hartmann 2002] where 176 brain lesions were assessed, showed that the use of MRS for the diagnosis of brain lesions together with structural MRI, compared to structural MRI alone, allowed to increase by 15.4% the number of correct diagnoses, decrease 6.2% the diagnostic errors and reduce in 16% the number of equivocal cases.

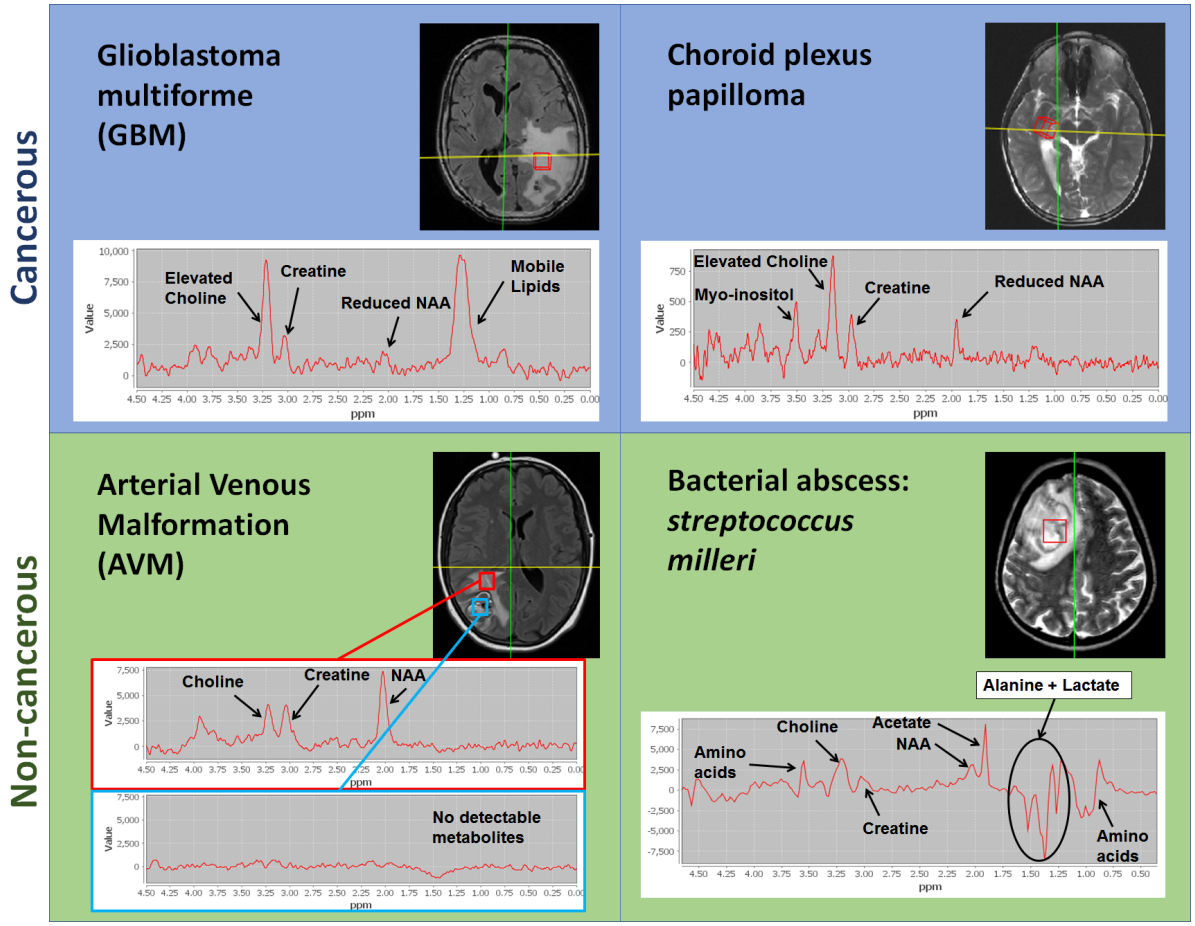

Figure 4.3: Four different types of lesions and corresponding spectra. All spectra acquired at 1.5T and 3.0T using PRESS, TE=135 ms. The axial images show the location of the spectra. For the bacterial abscess case, given that not all metabolites were included in the available basis set, peak assignment was made based on [Lai 2005]. The data shown in this figure was acquired at Inselspital, Bern.

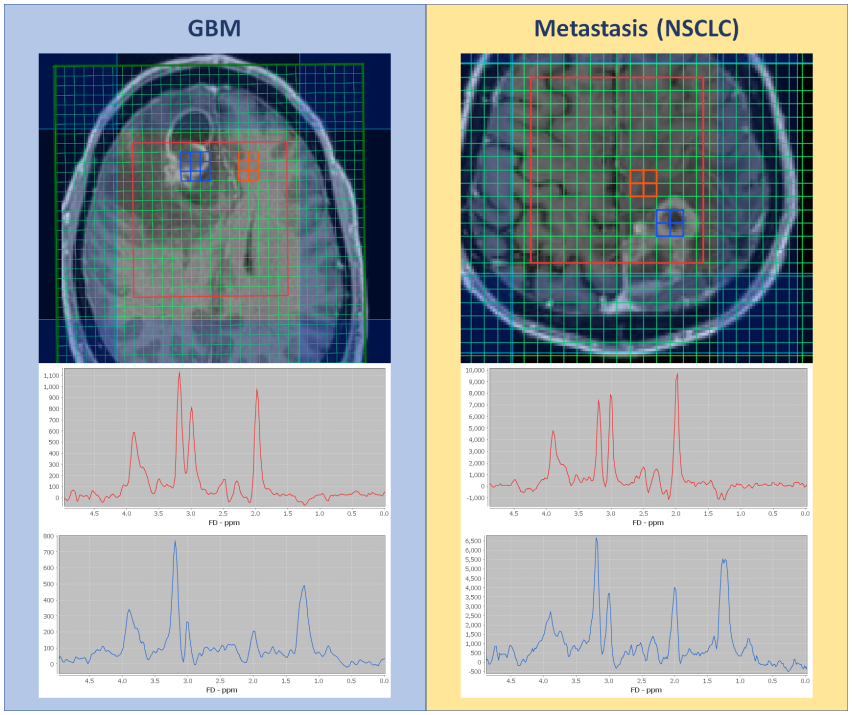

Figure 4.4: The figure shows an example comparing the spectra of the contrast-enhanced tumor region (blue region and spectrum) and of the peritumoral region (red region and spectrum) between GBM and metastasis. The primary tumor of the metastasis case is non-small cell lung cancer (NSCLC). Spectroscopy acquired at 1.5T using PRESS (TE=135ms). The aim of this figure is to present an example of what is described by [Server 2010] and [Baumert 2006] regarding the higher Cho/NAA and Cho/Cr of the peritumoral region of GBM tumors when compared to metastasis. The data shown in this figure was acquired at Inselspital, Bern.

4.2.2 Metastasis vs high-grade tumor

The distinction between brain metastases and high-grade tumors, such as glioblastoma, is one of the biggest challenges in initial diagnostics of brain tumors. MR spectra of both tumor groups may present very similar spectra which are difficult to distinguish by simple visual inspection. Despite this, a few studies have looked at the use of systematic approaches that use information derived from MRS and are able to distinguish metastases from high-grade tumors with reasonable accuracy.

A study from 2009 [Server 2010], that analyzed 73 patients with histologically confirmed diagnosis (53 high-grade tumors, 20 mestastases) looked at the peritumoral region to distinguish these two types of tumors. The results of the study showed that by defining a simple threshold on Cho/NAA applied to peritumoral spectra, a sensitivity of 100% and specificity of 91.1% could be achieved. Another study from Wijnen et al. [Wijnen 2012], confirmed that the peritumoral region contains relevant information for the distinction between metastases and high-grade tumors. The general hypothesis for the observed results is that high-grade tumors, namely GBM, are more infiltrative than most of the brain metastases. Consequently, the detection of elevated tumor markers in the region surrounding the tumor enables the distinction between these two tumor types (see example Figure 4.4). However, depending on the primary tumor, mestastases may present more or less infiltrative phenotype [Baumert 2006] which may complicate the differentiation between high-grade tumors and mestastases.

In this thesis, chapter 8 explores a method named *convex-Non-negative Matrix Factorization* (cNMF described in section 3.2.2) to highlight differences between MRSI data of GBM and metastasis patients.

4.2.3 Low-grade vs high-grade tumor

Several studies have confirmed the ability of MRS to grade brain tumors, namely astrocytomas, as discussed in the review paper by Hollingworth et al. [Hollingworth 2006]. As an example, a study from 2003 [Herminghaus 2003] used a linear discriminant analysis (LDA) classifier trained using 126 cases (42 WHO grade I/II, 30 WHO grade III and 54 WHO grade IV). The classifier was tested in a dataset of 94 new cases, composed of 30 grade I/II, 29 grade III and 31 grade IV. Overall the method classified correctly 86% of the cases, and differentiated low grade (grade I/II) from high-grade tumors (grade III/IV) with a success rate of 95%. As is described in the paper from Herminghaus et al., the ability to perform tumor grading based on MR spectra is related to the relationship between malignant transformation and several metabolites visible with MRS. In the malignant transformation from a low to a high grade glioma tumor, the first observed change is an increase of choline compounds due to the increased membrane turnover of the tumor cells, accompanied by the appearance of lactate due to anaerobic glycolysis. Simultaneously, tumor proliferation leads to the destruction of neurons, which is visible in MRS as a decrease in NAA. Consequently, tumor grade correlates positively with Cho/NAA and Cho/Cr, as shown in [Naser 2016].

4.2.4 Tumor-progression vs radiation necrosis

Radiation therapy leads to the focal destruction of tumor and healthy brain tissue (radiation necrosis) which can lead to the disruption of the blood-brain-barrier (BBB). Consequently, in T_{1c} sequences, new contrast enhancement is observed in those regions. Given that disruption of the BBB and contrast enhancement are also observed in high-grade tumors, radiation necrosis and

tumor progression are often difficult to discriminate in T_{1c} and other structural MR sequences. The correct differentiation between radiation necrosis and tumor progression is essential for the follow-up evaluation of brain tumor patients.

Despite the similarities in MR imaging, MR spectra of progressive tumor tissue and radiation necrosis can be distinguished. Given that tumor proliferation leads to an increase in free choline concentration, tumor progression is characterized by significantly higher Cho/Cr and Cho/NAA values, when compared to radiation necrosis [Weybright 2005]. As described by Weybright et al., the authors were able to discriminate correctly these two tissue types in 27 out of 28 cases, using only the ratios Cho/Cr and Cho/NAA. Similar results were obtained by Rock et al. [Rock 2004], showing that MRS provides relevant information for the distinction between radiation necrosis and tumor progression. Moreover, as referred by Rock and his colleagues, the combination of MRS with other modalities such as Apparent Diffusion Coefficient (ADC) may further improve the discrimination performance.

4.2.5 Tumor border delineation

Conventional brain tumor segmentation uses as reference structural images such as T_1 post contrast (T_{1C}), T_2 and FLAIR to identify solid tumor and edema. The solid tumor contours are normally defined as the region delimited by contrast-enhancement in T_{1C}, and the edema as the region that surrounds the solid tumor and which appears hyperintense on FLAIR images or T_2. Definition of tumor contours is essential for tumor patient management, namely for the assessment of response to treatment, for radio-therapy planning and for guiding surgery. Despite its importance, the abnormalities seen in structural images significantly underestimate the extent of brain tumors such as glioblastoma. A study from Yamahara et al. [Yamahara 2010] that analyzed 7 autopsy brains of deceased glioblastoma patients detected dense tumor cell infiltration at a distance from up to 14 mm of the tumor border defined in T_{1C} images. Moreover, tumor infiltration was found in regions with normal appearance in T_2 images. Similarly, Pallud et al. [Pallud 2010] showed that in diffuse low-grade oligodendrogliomas tumor cells can be found from up to 26 mm from MRI defined abnormalities.

In order to address this issue, some authors have turned to MRSI as a solution to improve brain tumor delineation. Stadlbauer and his colleagues [Stadlbauer 2004] showed that Cho/NAA maps can better estimate glioblastoma tumor extent and that the region of abnormal Cho/NAA values extends beyond the region defined by T_2 abnormalities. In this study, all biopsy samples collected from regions normal in T_2 but abnormal in Cho/NAA maps showed tumor infiltration. A similar study [Cordova 2016] using a higher spatial resolution MRSI sequence, showed that Cho/NAA positively correlates with tumor cell density. Moreover, in the same study, Cho/NAA maps were shown to predict contrast-enhancement from 4 to 5 months before its appearance on T_{1C} images. Correlation between Cho/NAA values and tumor cell density had previously been shown by Ganslandt et al. [Ganslandt 2005]. In the same study, it was once more demonstrated that the region defined by the Cho/NAA maps exceed the tumor affected area defined in T_2. This was observed in all of the 7 patients analyzed.

In parallel with these positive results, a phase II trial [Einstein 2012] was conducted to evaluate the use of Cho/NAA maps for planning radiotherapy. The study showed a significant increase in patient survival after incorporation of MRSI for defining target volumes.

Overall, the different results of the literature here described demonstrate the added value of MRSI for improving brain tumor segmentation. However, given that MRSI, namely high spatial resolution sequences and other more experimental approaches used in research, is less

available than conventional structural images such as T_{1C} and T_2, MRSI is not yet part of the standard procedure for brain tumor segmentation.

In this thesis, chapters 10 and 11 present novel approaches that explore MRSI data with the aim of enabling better delineation of brain tumors.

Part II

Quality management of MRS and automation of quality control

5 | Quality Management in *in vivo* proton MRS

This chapter is based on the following publication:

N. Pedrosa de Barros, J. Slotboom, *Quality management in in vivo proton MRS*, Anal. Biochem. 529 (2017). doi:10.1016/j.ab.2017.01.017.

5.1 Abstract

The quality of MR-Spectroscopy data can easily be affected in *in vivo* applications. Several factors may produce signal artefacts, and often these are not easily detected, not even by experienced spectroscopists. Reliable and reproducible *in vivo* MRS-data requires the definition of quality requirements and goals, implementation of measures to guarantee quality standards, regular control of data quality, and a continuous search for quality improvement.

The first part of this review includes a general introduction to different aspects of quality management in MRS. It is followed by the description of a series of tests and phantoms that can be used to *assure* the quality of the MR system. In the third part, several methods and strategies used for *quality control* of the spectroscopy data are presented. This review concludes with a reference to a few interesting techniques and aspects that may help to further improve the quality of *in vivo* MR-spectra.

5.2 Introduction

Quality management has four major components [Evans 1999]:

1. *Quality planning*;
2. *Quality assurance*;
3. *Quality control*;
4. *Quality improvement*;

Quality planning consists in the definition of quality requirements and goals. *Quality assurance* deals with the prediction of possible errors and defines a set of measures to prevent them. *Quality control*, on the other hand, deals with situations where errors are created and its main aim is to detect them and avoid that these cause further complications. Finally, *quality improvement* defines how the quality of the results can be further improved. In Figure 5.1, a few examples of different tasks that fall in each of the different quality management components in MR spectroscopy are given.

Quality management is required to ensure data reproduction, and is especially important in applications such as individual patient diagnostics, and multi-centre studies where data

Figure 5.1: Exemplary tasks that belong to the different components of MRS quality management.

acquired in different sites, and frequently with different measurement conditions, is combined [Podo 1995, van der Graaf 2008]. In multi-centre studies, apart from the quality management performed at the level of each centre, further measures need to be implemented to ensure data compatibility.

It is important to understand that *none* of the four quality management components are closed or independent. Tests performed in *quality assurance* can loosen or tighten quality requirements previously defined in *quality planning*, or even define new targets for *quality improvement*. Moreover, information collected in *quality control* may lead to new tests and prevention measures in *quality assurance*. And besides this, some actions may belong to more than one component. For example, as proposed by Milan Hájek *et al.* in [Hájek 2000], quantification calibration factors can be used both for *quality assurance*, improving the accuracy of the measurement, as well as for *quality control*, helping to detect relevant changes in MR-scanner performance.

The four components of quality management will now be discussed in more detail.

5.2.1 Quality planning in MRS

Quality planning in *in vivo* MR-spectroscopy involves the definition of quality targets and should give answers to questions such as: what is the minimum concentration and voxel size that needs to be detectable? What is the coarsest MRSI resolution still acceptable? What are the minimum levels of SNR and FWHM that are considered acceptable? What is the maximum acquisition time that is still acceptable to ensure patient or test person wellbeing? How much error is tolerated in absolute quantification? What is the maximum acceptable rate of rejected spectra in *quality control*?

Quality planning is, obviously, highly dependent on the application. For instance, in certain diseases it may be required to detect a metabolite in very low concentrations (e.g. 2-hydroxy-

glutarate in gliomas [Dang 2009, Choi 2012]). On the other hand, depending on the size of the structure that is analysed, different resolution levels may be needed. Moreover, depending on whether the application is clinical or not, the maximum possible acquisition time can change considerably. Given that most of these factors are interdependent, some compromises need to be made. Therefore, *quality planning* should start by the definition of the main goals and of their relative importance.

5.2.2 Quality assurance in MRS

Once the goals and quality requirements are set, *quality assurance* actions may be initiated. *Quality assurance* starts with the prediction of what might fail, and with identification of the factors that may lead to failure. Once these factors are identified, an adequate battery of tests should be defined. These tests should provide clear knowledge of the limitations of the scanner hardware and software (pulse sequence). Whenever possible, the hardware and pulse sequence should be adjusted to increase the robustness to the factors identified. *Quality assurance* is often an iterative process, where tests and corrections are repeatedly performed until the targets defined in *quality planning* are met. *Quality assurance* in MRS includes the following actions:

1. Improving the robustness of the sequence and scanner;
2. Providing information regarding the conditions under which the quality of the spectra is guaranteed (e.g. manuals stating how the MRS measurement should be planned and the data processed);
3. Providing information regarding the limitations of the pulse sequence and scanner (should be provided by the manufacturer in the system specification);

In section 5.3 a few phantoms and tests to assure the quality of the spectra are presented.

5.2.3 Quality control in MRS

As stated above, the aim of *quality control* is to detect errors, and to prevent the impact that these can have in the obtained results. Consequently, the definition of the appropriate control strategy starts with the identification of all critical parts of the *acquisition* and *processing* of the data: what aspects are most likely to fail and what factors have the biggest negative impact on the quality of the measurement (e.g. failing auto-shim)?

After identifying the critical points, a strategy for detecting anomalies (e.g. unexpected sudden phase jumps during acquisition or bad water suppression) can be defined. It is important to analyse which information needs to be collected for an accurate anomaly detection. If needed, changes can be made to the protocol so that the required information for *quality control* is collected. The collected information should grant *traceability* to the system, or in other words, should provide enough information for determining *what failed*, *when* and *how*.

Once anomaly detection is implemented, the set of actions that should follow the different outputs of the detector, need to be defined. In some cases, the result can be the simple rejection of data (quality filtering), whereas in others certain pre-processing steps may be triggered. An important outcome of the *quality control* is the information regarding the *reliability* of the measurement. A common example of this is the CRLB, which gives information regarding the spectral reliability, i.e. the lowest possible errors in the estimated quantification parameters.

In *in vivo* spectroscopy, *quality control* focuses on mainly two aspects: signal quality, and quantification quality. Signal *quality control* aims at detecting signal artefacts and bad quality features, and often leads to the rejection of data or the identification of those cases that are likely to produce unreliable results. Quantification *quality control* deals with the performance of the quantification process and provides measures of reliability of the output. These measures can be used for quality filtering.

Section 5.4 describes different methods for *quality control*.

5.2.4 Quality improvement in MRS

Finally, *quality improvement* starts with the knowledge of the performance of the current system and aims at improving it. The improvements go beyond the minimum requirements and goals defined in *quality planning*, and are not necessarily triggered by undesirable events or failures.

Examples of common measures used for *quality improvement* in MRS include the development and implementation of new pulse sequences, or the improvement of the currently used ones; the use of better RF-coils; a new scanner or scanner upgrade; improved education of MRS-users; improvements in data post processing routines; etc.

In section 5.5, several aspects that should be considered for quality improvement will be presented.

5.3 Tests and phantoms for quality assurance

The discussion on MRS *quality assurance* and phantom testing cannot avoid mentioning the European concerted action on quality assessment in *in vivo* NMR spectroscopy [Podo 1995, Bovée 1995, Leach 1995, Keevil 1995, Howe 1995, de Beer 1992] from 1995. Even though it was performed more than 20 years ago, this action was an important step for the establishment of MRS, and the published material is still up-to-date. This action was motivated by the need for standardized MRS-procedures. It included test measurements and methods for data analysis, using appropriate test objects and substances to determine quality measures. Such tests are essential for the standardization of the technique, as well as to allow data sharing between different centres, and realization multi-centre studies on large patient numbers.

Another concerted action followed, with focus on absolute metabolite quantification by in vivo NMR spectroscopy [Podo 1998, Keevil 1998, De Beer 1998a, Bovée 1998, De Beer 1998b]. In the context of that action, a method to evaluate the localization performance in MRSI [Bovée 1998] was proposed which is described in more detail in subsection 5.3.4.

This section covers several aspects of phantom testing, starting with some of the key findings of the 1995 concerted action.

5.3.1 Two-compartment phantoms for SVS sequences

Leach *et al.* discussed in their work [Leach 1995] a few practical considerations made in the design, construction and solutions used in two MRS phantoms of the 1995 concerted action. One phantom, the so called "Spectroscopy Test Object 1" (STO1 – Figure 5.2), was a sphere with an inner cubic compartment, and was aimed at testing *single voxel localisation* methods. The second phantom, named "Spectroscopy Test Object 2" (STO2 – Figure 5.3), was a cylinder with a moveable cylindrical inner container, and was designed to test *slice localisation* methods.

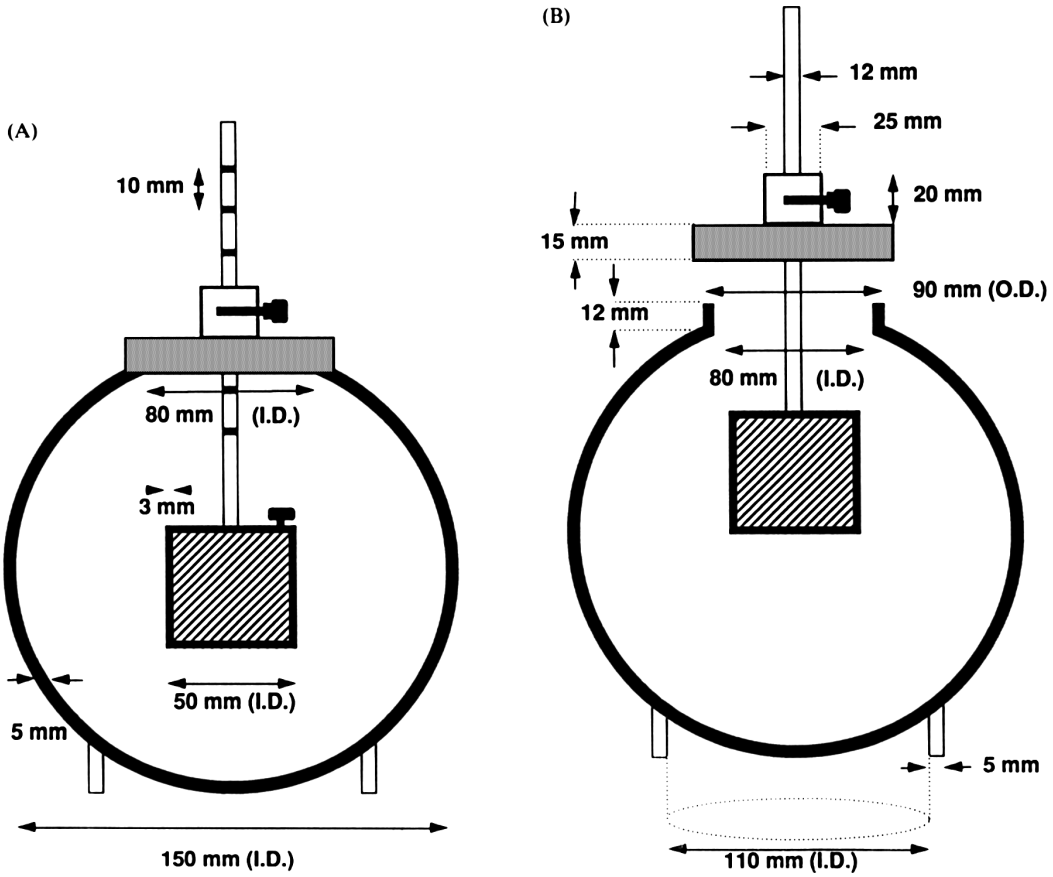

Figure 5.2: Spectroscopy Test Object 1 (STO1) of the 1995 Concerted Action. (A) and (B) show different perspectives and dimensions of the object. Source of the image: [Bovée 1995]. Copyright by Elseviser Science Ltd.

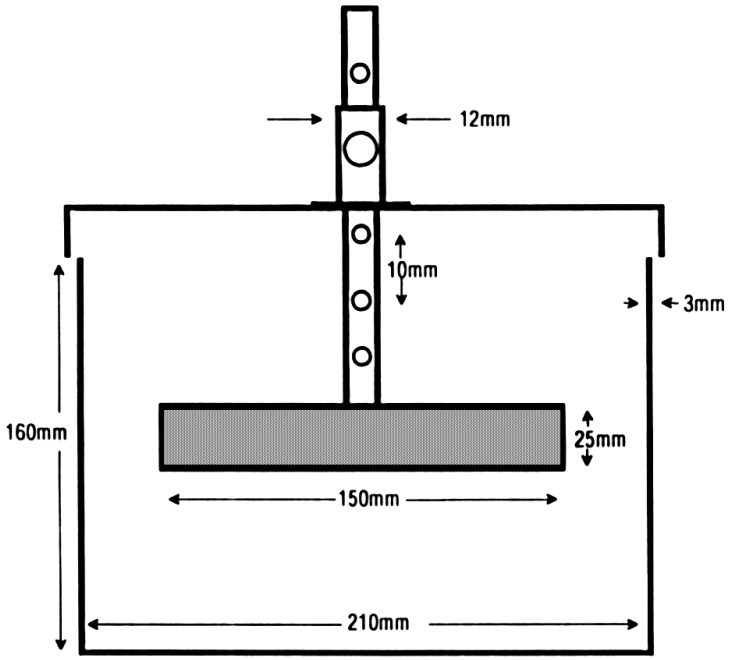

Figure 5.3: Spectroscopy Test Object 2 (STO2) of the 1995 Concerted Action. Source of the image: [Bovée 1995]. Copyright by Elseviser Science Ltd.

Table 5.1: Susceptibility difference with water for several materials. Data source: [Bovée 1995].

Material	**Susceptibility field shift (ppm)**
Perspex	<0.005
Polycarbonate	0.014
Glass	0.023
Polyethylene	0.087
Teflon	0.096
PVC	0.12

Both phantoms and procedures were designed for testing the localization performance of pulse sequences using homogenous bird-cage coils.

One of the critical parameters evaluated before construction of the phantoms was the susceptibility difference of several materials with water. PVC, Perspex, Teflon, Polycarbonate, Polyethylene and Glass were tested, with Perspex showing the smallest susceptibility difference to water and consequently becoming the material of choice to build the phantoms (Table 5.1). Smaller differences in susceptibility pose fewer challenges for shimming and, consequently, guarantee higher field homogeneity over the excited volume. In Figure 5.4 we show a Susceptibility Weighted Image (SWI) of a phantom made using an unknown plastic material. The figure demonstrates that the chosen material shows noticeable susceptibility differences with the surrounding solutions. Additionally, the phantom also shows the signal degrading off effects of paramagnetic particles contained in the material and air bubbles. This example also shows that these susceptibility artefacts affect the SNR and linewidth (measured by the damping factor used in quantification) of the spectra. As it can be seen, the lower right cube is the most affected, whereas the top left cube shows better quality spectra as well as less susceptibility artefacts. Another explanation for the poor quality of the spectra might be the cubic geometry of the

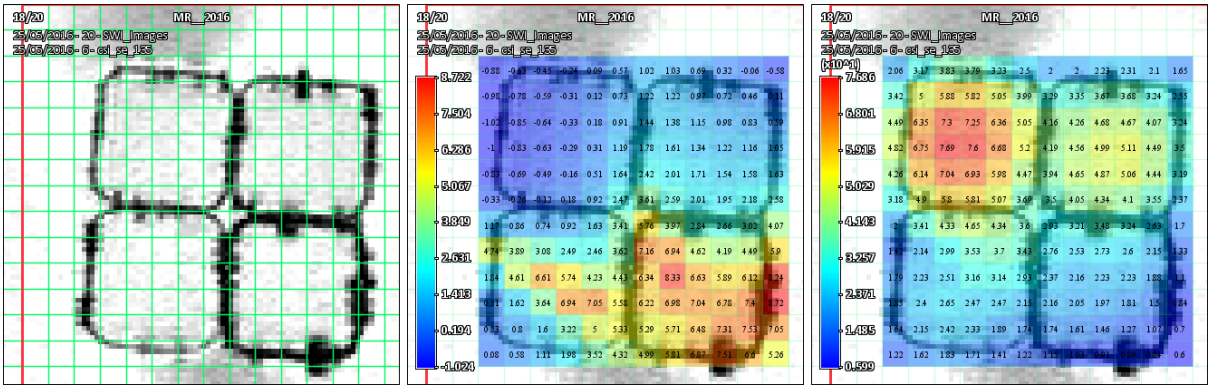

Figure 5.4: Susceptibility artefacts caused by an unknown plastic material used for the phantom construction. The phantom is made of 4 cubes of the same plastic material. Each cube contains a solution with an identical NAA concentration of 70 mM. The outer compartment was filled with water. From left to right we show a susceptibility weighted image (Left); the NAA damping factor in Hz (comparable to linewidth) used in QUEST (Middle); and SNR of NAA (Right). The measurement was performed using a 2D MRSI PRESS sequence in a 1.5 Siemens Aera Scanner (TE 135ms and TR 1500ms). The centre of the 4 cubes was placed at the isocentre of the magnet.

inner compartments, which is not optimal for obtaining good shimming results.

Regarding the solutions used in phantoms STO1 and STO2, the outer volume was filled with demineralised water with 0.1 mM Gd-DTPA, 0.05 mM $MnCl_2$ and 0.09 M NaCl (T_1 of 814 ms and T_2 of 235 ms). For the inner compartment the authors used Polydimethylsiloxane, and optimized its viscosity to achieve adequate T_1 and T_2 values (viscosity 1070 mPa.s, T_1 740 ms and T_2 180 ms). Alternatively, a conventional metabolite solution could have been used.

From the same concerted action [Podo 1995, Bovée 1995, Leach 1995, Keevil 1995, Howe 1995, de Beer 1992], another paper [Bovée 1995] was published by Bovée and his colleagues on quality assessment using the two described phantoms (STO1 and STO2). The paper introduces a set of NMR signal related parameters that describe different quality aspects that can be evaluated using those phantoms:

1. *Linearity*: Relation between signal amplitude and excited volume (VOI) and between the signal amplitude and concentration. The relation with concentration can be analysed simply by performing several measurements of the same excited VOI, with varying concentrations. Relation with volume can be analysed by changing the VOI. Ideally, both relations should be linear.

2. *Chemical Shift*: Analysis of the chemical shift difference between the inner and outer compartment peaks depending on the localization method. In this test, the difference between peaks is compared between localized and non-localized spectra. The localized spectra are acquired with a VOI centred in the inner compartment. This test may help to detect B_0 inhomogeneities and eddy currents.

3. *Spectral resolution: FWHM* (full width at half maximum of a certain resonance) and full width at $1/10^{th}$ of maximum height measured for both peaks with and without localization

4. *SNR* per unit concentration in fixed volume.

5. *Selection efficiency*: ratio between the signal strength obtained from the VOI using the localization method, divided by the maximum possible signal from the VOI measured by

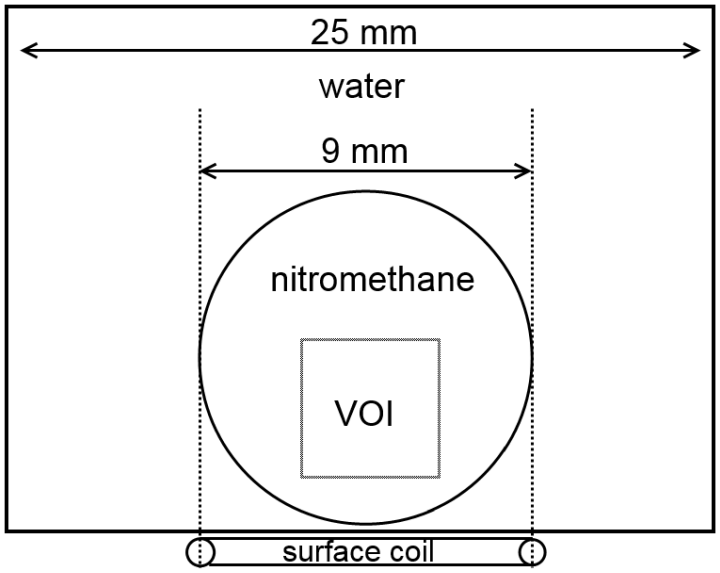

Figure 5.5: Two-compartment phantom used in the work of Slotboom *et al.* [Slotboom 1991, Slotboom 1993]. The phantom was designed to test localization pulse sequences when using surface coils at high B_0- field-strength. Image source: [Slotboom 1993].

a non-localized pulse sequence. Sources of inefficiency include, for example, excitation imperfections, eddy currents or chemical shift displacement artefacts.

6. VOI profile and position accuracy.

7. *Suppression factor*: defined as the ratio A/B where A and B are the unwanted signal strengths measured, respectively, without and with suppression.

8. *Signal stability*: repetition of the non-localized experiment several times and determination of the mean and standard deviation of the signals.

Detailed information on how to perform the different measurements and tests can be found in the original paper [Bovée 1995], and an example of an application of the assessment protocol can be found in the paper of the multicentre study [Keevil 1995] of the same concerted action. In that multicentre study, the performance of PRESS and STEAM in proton MRS were compared using data collected from 8 different centres. The comparison confirmed what is now commonly known: higher selection efficiency was obtained using PRESS due to the implicit 50% signal loss of STEAM, and lower signal contamination was obtained using STEAM due to higher selectiveness of the 90 degree pulses used in STEAM compared to the 180 degree pulses used in PRESS.

Another two-compartment phantom containing water and nitromethane (Figure 5.5), was used by Slotboom *et al.* [Slotboom 1993, Slotboom 1991] to evaluate the localization performance, at 7T, of SADLOVE (later LASER) against PRESS and STEAM using surface coils. The analysis uses the same principles as in [Leach 1995] to evaluate localization efficiency. The following experiments were performed to compare selection performance between sequences:

1. Non-localized FID experiment using 90° adiabatic pulse with resulting peak-areas W_1 and N_1 for respectively water and nitromethane.

2. Localization experiment with selection gradients switched off during RF-pulse application. Resulting peak areas for water and nitromethane: W_2 and N_2.

3. Localization experiment with selection gradients on. Selection of a 4x4x4 mm cubic VOI in the nitromethane container, as seen in Figure 5.5. Resulting peak areas (water and nitromethane): W_3 and N_3.

Using the mentioned peak-areas the following parameters were derived:

1. Suppression efficiency: W_1/W_3;
2. Signal efficiency: $(W_2+N_2)/ (W_1+N_1)$;
3. Overall performance: Signal efficiency / Suppression efficiency;

The study evaluated the performance of the three sequences as a function of the amplitude of the selection pulses. For the test values obtained using surface coils, SADLOVE showed overall performance values that were from 3 to 12 times higher than the ones of STEAM and PRESS. STEAM and PRESS showed similar results, with STEAM having higher suppression efficiency and PRESS higher signal efficiency, in accordance with the results of the multicentre study described above.

When comparing the phantom of Figure 5.5 with STO1 and STO2, it is noticeable the size difference. The difference is related with the size of the bore used: STO1 and STO2 were designed for use in clinical scanners whereas the nitromethane and water phantom was used in a small-bore animal scanner. The sizing aspect is also discussed in another paper [Howe 1995] of the 1995 concerted action. In that work the authors successfully applied the same testing protocol as defined by Bovée *et al.*in [Bovée 1995] to a small bore system but had to perform some changes in the size and wall thickness of the phantoms.

A more recent example of *quality assurance* using phantom-tests comes from another European multi-centre project: INTERPRET [Pérez-Ruiz 2010, Tate 2006, Julià-Sapé 2006]. The aim of this project was to build automatic spectra classifiers for brain tumour typing based on proton SVS spectra. As mentioned in the introduction, *quality assurance* is a key aspect of multi-centre projects, given the need to assure data compatibility between the participating centres. In INTERPRET, data compatibility was assured by the protocols [van der Graaf 2008] defined for *quality assurance*, which will be described below, and for *quality control*, which will be described in the *signal quality control* section (5.4.1).

INTERPRET used similar phantoms and measurements as the ones described in the concerted action of 1995. The two-compartment Perspex phantom used had a cylinder shape and contained a smaller inner cubic compartment with a 20mm diameter and 3 mm wall thickness. The outer compartment was filled with 0.15M sodium acetate solution (pH 8.3) and the inner compartment with a solution of 0.1M lithium lactate and 0.1M creatine (pH 7.6). Sodium azide (1g/L) was added to avoid bacterial growth. Three different measurement protocols were used for testing the MR-scanners from the different centres. *Protocol A:* a simple protocol that was repeated every 2 months using the INTERPRET phantom. *Protocol B:* a more extensive one using the same phantom and performed every year, and after each hardware or software upgrade. *Protocol C*: performed *in vivo*, applying the same measures as *protocols A* and *B* in a group of at least five healthy volunteers. *Protocol A* consisted in three SVS measurements with a cubic VOI of $(15\ mm)^3$: one VOI was placed exactly in the centre, a second VOI placed 40 mm below the central VOI, and a third VOI placed 40 mm left of the central VOI. In *protocol B*, VOIs with increasing size centred in the inner compartment (first contained inside the inner cube and then including the inner cube and the surrounding volume) and placed 40 mm below the centre

of the cube, were used to test volume selectivity and signal linearity. Long TE (136ms) SVS PRESS and short TE SVS STEAM (20ms) and PRESS (30ms) were used in the measurements.

The results obtained using the bimonthly measurements (*protocol A*) allowed to detect a localisation problem due to a faulty RF-transmitter in one of the participating sites. The fault was detected based on an unexpected increase of the water bandwidth with no change in SNR. Apart from this occurrence, the SNR was above 50 and the water bandwidth below 4 Hz for every site in all performed tests. The results of the *protocol B* showed that the *signal contamination* – ratio between outer volume signal and total signal – measured with a voxel of $(20\ \text{mm})^3$ exactly covering the inner compartment of the phantom varied between 0 and 5.2%. The *signal efficiency* test parameter – which is the ratio between the inner volume signal for a certain VOI and the maximum inner volume signal measured by a non-localized experiment – was determined. The observed *signal efficiency* ranged between 71 and 75% for a VOI with the dimensions of the inner compartment, and between 90 and 98% for a slightly larger VOI of $(25\ \text{mm})^3$. Finally, the spectra acquired in the *in-vivo* tests using healthy volunteers (*protocol C*) showed satisfactory quality, with no artefacts, high SNR and narrow linewidth.

5.3.2 Creating compartments without walls

Given that walls in phantoms may introduce susceptibility differences and other problems, some groups have proposed some interesting alternative phantoms that create different compartments without the use of a separating material. One of such examples is the oil and water phantom proposed by the Burtscher *et al.* [Burtscher 1999]. The phantom is a Perspex sphere of 15 cm diameter filled half way up with a water solution containing 0.05 mM $MnCl_2$, 0.2 mM gadolinium DTPA and 0.09 M NaCl. The other half was filled with a polydimethylsiloxane (PDMS) oil with a viscosity of 1000 mPa.s. The relaxation properties were: $T_1 = 810$ ms and $T_2 = 220$ ms for the aqueous solution, and $T_1 = 770$ ms and $T_2 = 200$ ms for the oil. The two substances had a chemical shift difference of 4.5 ppm. The phantom was used to compare the localization performance of a spin echo (SE) and STEAM sequences with focus on unwanted lipid signal contamination. Different measurements were made with different voxel sizes that were moved along the interface between the two solutions. Short TE (10 ms) STEAM showed smaller levels of contamination, when compared with long TE (270 and 135 ms) SE and STEAM. This is explained by the larger bandwidth (2560 Hz) of the pulses used in short TE STEAM when compared to the pulses used in the other sequences (850 Hz) – relation between pulse bandwidth and localization accuracy discussed in section 5.5.4. In STEAM, bigger voxel sizes had significant lower levels of contamination than smaller voxels. This effect was explained by the reduction in the quality of the slice profile for smaller slice thicknesses, and the Hamming filter applied to the RF-pulse that broadens the excited slice and which has a more significant effect in smaller widths. The relation between contamination and VOI size was not observed for the SE sequence.

Another example of a phantom without inner walls was proposed by Heikal *et al.* in 2008 [Heikal 2008]. The phantom was used to evaluate the delineation accuracy of 2D-PRESS, 2-echo Spin Echo Spectroscopic Imaging (SESI) and 4-echo SESI. The authors used porcine gel to create the two different compartments. For the construction, a cylinder was filled with solution with 3 mM choline chloride, 10mM creatine hydrate and 5% weight porcine gel. An acrylic wedge was inserted in the solution to create the space for the inner compartment. After solidification of the first solution, the wedge was removed and the corresponding space filled with a similar solution as the previous one but with 10 mM choline chloride. 2D PRESS showed the best

results, allowing for a more accurate delineation of the inner compartment.

5.3.3 2D-Array phantom for testing MRSI sequences

Woo *et al.* proposed in 2006 a phantom for testing MRSI sequences using a 2D-array of cone-shape vials [Woo 2007]. The 64 vials (arranged in an 8x8 matrix) were made of polyethylene and contained a rubber lid for sealing. The different vials were fixed inside a cylindrical phantom body made of acrylic resin. Two internal circular plates were used to fix the phantoms (one contained equally spaced holes displayed in an 8x8 matrix, and the other a regular quadrilateral hole). The outer compartment was filled with a 0.7 g/L copper sulphate solution (Cu_2SO_4). The vials were filled with different solutions using the following substances: NAA, creatine, choline, lactate, alanine, myo-inositol, GABA, glutamine and gadopentetic acid. Sodium azide (0.01%) was added to these solutions to prevent bacterial growth. All vials contained also 56 mM NaOH and 50 mM KH_2PO_4. Using the phantom, a 2D PRESS (TE/TR = 40/1500 ms) with 20x20 spatial encodings was tested. Relation between amplitude and volume was analysed by placing the MRSI at different levels or heights of the conic vials. Even though this phantom provides a flexible solution for testing MRSI sequences with different metabolite concentrations, the phantom design might create some complications in terms of shimming. This may explain the several poor quality phantom spectra presented in the paper. Multi-compartment grids can provide interesting test objects for Spectroscopic Imaging sequences, however, as was described above, special attention should be given to the materials and design of the phantoms, given that spectroscopic sequences are highly sensitive to field inhomogeneities. Phantom geometries, such as large rectangular shapes, can give rise to B_0-inhomogeneities that cannot be globally shimmed by the available shim set. The signal artefacts might have been lower with a simplified phantom design and the use of materials with smaller susceptibility differences with water, such as Perspex, polycarbonate or glass. A grid of gelatine[Heikal 2008] cubes with different metabolite concentrations could potentially lead to better results.

5.3.4 Point-Spread-Function analysis

The Point-Spread-Function (PSF) is the response of the image or in this case, MRS-image, to a source with the exact size of one voxel (a so-called Dirac δ function). The PSF of one voxel shows the contribution to the signal of the neighbouring voxels. Detailed information on how to determine and analyse the PSF can be found in the work of Bovée *et al.* [Bovée 1998] from 1998. In this paper, the authors suggest to perform localization *quality assurance* in MRSI by comparing the theoretical PSF against the experimental one. It was hypothesized in that work that deviations between the two might be caused by eddy currents and field inhomogeneities. In order to measure the experimental PSF the author suggests the use of a simple narrow test tube filled with water, with an inner diameter not larger than half the width of the non-interpolated pixel. More detailed information on how to perform the proposed analysis can be found in the original paper.

5.3.5 3D-printed phantoms

Following the recent progress in 3D printing techniques, a few authors have evaluated the use of this technique in phantom construction. 3D-printing provides an enormous freedom in terms of phantom design and it is becoming increasingly available for industrial use as well as for the general consumer.

In order to evaluate the feasibility of using 3D-printed phantoms, Bieniosek *et al.* [Bieniosek 2015] compared a commercial phantom with the 3D-printed version of it. The shape was cylindrical, with several cylindrical inner-compartments of different sizes. The phantoms were evaluated for CT and PET and MRI, and showed that the differences in the results obtained between the original phantom and the 3D-printed version were within the measurement uncertainties of the setups. Unfortunately, the phantom was not tested for MRS.

Another example is the 3D-printed Shepp-Logan proposed by Kasten *et al.* [Kasten 2016]. The Shepp-Logan phantom [Shepp 1974] is a numerical phantom proposed in 1974 for evaluation of reconstruction techniques. This standard test image contains different ellipsoids of different shapes and intensities, which makes it an interesting object for benchmarking different reconstruction methods. Besides its general use, to the best of our knowledge, the phantom existed only *in silico*. Kasten and his colleagues created a 3D model of this numerical phantom, which was then 3D-printed, that allows easily changing the solutions of inner compartments. The different compartments of the phantom were filled using different concentrations of NAA, creatine, choline, and sodium lactate. Corn oil was added to the peripheral compartment, simulating the lipids of the scull and scalp. The work of Kasten *et al.* showed that the 3D-printed Shepp-Logan phantom is an interesting test object for evaluating MRSI reconstruction techniques.

5.3.6 Check-list for phantom design and construction

The following list gives an overview of several important points that must be considered in phantom design and construction.

Shape and design:

1. Is the chosen shape easily shimmed?
2. Are the dimensions (including material thickness) adequate for the bore size?
3. How many compartments are needed?
4. Are there viable solutions to avoid internal walls? If not, can the walls be made thinner?
5. Can it be well sealed?
6. Does it allow changing the solutions of the different compartments easily?

Materials:

1. What is the susceptibility difference with water?
2. Does it react with the solutions we want to use?
3. How resistant is the material?

Metabolite solutions:

1. What are the relaxation properties? Are these similar to those of the target tissue?
2. What is the chemical shift? Is this close to the metabolites we want to measure *in vivo*?
3. How is the chemical shift of the substances affected by temperature and pH?

4. Are the spins of the chosen substances J-coupled or uncoupled?
5. Are the compounds chemically stable?
6. Do they have special storage requirements (e.g. need to be kept refrigerated)?
7. Do they react with the wall materials of the phantom?
8. Are the solutions dangerous? (Explosive, Corrosive, *etc.*)
9. Should a preservative be added to the solution? (e.g. sodium azide to prevent bacterial growth)

5.4 Methods for quality control of MRS data

In quantitative MRS, errors can occur during the signal acquisition phase, as well as during the signal processing phase. Consequently, *quality control* in MRS focuses mainly on these aspects:

1. Signal quality assessment and detection of signal artefacts after acquisition (SVS and MRSI);
2. Analysis of the quantification errors;

These two aspects are obviously related, but the approaches used in each one are different. In the next two subsections, we will cover some of the techniques that have been used for *signal quality control* and for *quantification quality control*.

Before diving into *quality control* methods, it is important to note that qualitatively good looking spectra (high SNR, narrow linewidths, no obvious artefacts) are not necessarily reliable [Slotboom 2009, Kreis 2004]. Several of the effects that can be analysed in *quality assurance*, such as the chemical-shift artefact, the signal contamination, localization inaccuracies, signal cancellation, etc. do not lead directly to clearly identifiable "bad-looking" signals nor even necessarily cause problems for the quantification. Consequently, these effects are not detected by the *quality control* methods that will be described in this section, methods whose input are simply the time and frequency domain signals.

5.4.1 Signal quality control

As already mentioned, *in vivo* MRS-signals can easily be affected by B_0- and B_1-field inhomogeneities, eddy currents, susceptibility differences, signal contamination (e.g. subcutaneous lipid signals), etc. In vivo proton spectra are particularly sensitive to the mentioned factors due to the relatively low SNR of the majority of the metabolites of interest compared to the SNR of water, which is around five orders of magnitude higher. As discussed in the previous sections, in the context of *quality assurance* some of the effects that the mentioned sources of artifacts can have are addressed, and measures for preventing them are defined. Still, signal artefacts are very common in *in vivo* spectra and *quality control* is an essential step in the processing pipeline. An overview over frequently occurring artefacts that can affect MRS signals, as well as some possible corrections, can be found in the work of Kreis from 2004 [Kreis 2004].

One of the major problems with MRS artefacts is their identification. Several artefacts are "silent" and do not show clear features that allow their easy detection (e.g. alterations in the macro molecular baseline). In addition, artifacts might also be simply camouflaged as "just

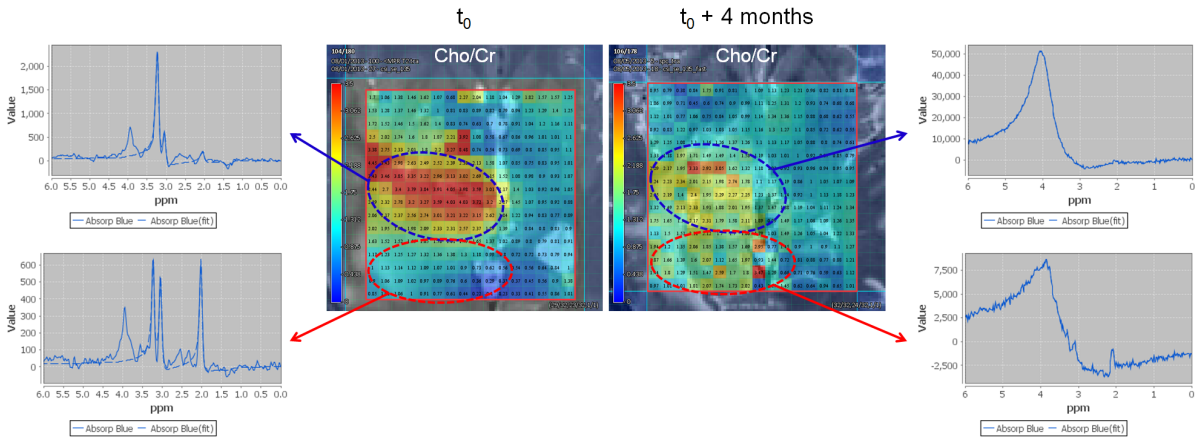

Figure 5.6: Example showing two Cho/Cr maps of a glioma patient acquired 4 months apart. Acquisition details: Siemens Aera 1.5 T, 2D-PRESS, 12x12 interpolated to 32x32, TE/TR 135/1500ms. The purpose of the example is to illustrate a situation where wrong clinical interpretations could be made from poor quality spectra, such as the spectra of the second acquisition. MRSI images should be checked for sufficient spectral quality *before* maps can clinically be interpreted. This quality check must be performed by a spectroscopist with thorough knowledge of the method, or an algorithm that can substitute this expertise.

another peak" (e.g. lipid signals excited by RF-pulse sidebands that are outside the intended excited volume VOI, and appear at wrong chemical shift position due to imperfect shimming). In clinical applications, for instance, this might lead to wrong conclusions drawn from artefact affected metabolic maps, since there is no time to investigate the quality of signal of all voxels of the MRSI-grid manually. To illustrate this, we included the example of Figure 5.6 where we present two Cho/Cr maps of the same patient and same localization acquired 4 months apart from each other. By looking at the change in Cho/Cr between those maps, one could identify two main regions: in blue a region with signs of response to treatment (reduction in Cho/Cr) and, in red, a region with signs of tumour progression (increase in Cho/Cr). It's only when we look at the average spectra over each region that we realize that the spectra of the second MRSI acquisition is highly affected by artefacts caused by surgical induced tissue changes (residual blood, metallic particles of surgical instruments), and that its quality is far out too poor to allow *any* clinical interpretation at all.

5.4.1.1 Rejection criteria

Several efforts have been made to define quality requirements in clinical spectroscopy. Worth mentioning here is the extensive checklist included in [Kreis 2004] and the rejection criteria defined in [Kreis 1993, Oz 2014]. According to these references, spectra should be rejected if:

1. FWHM > 0.1 ppm
2. Asymmetric line-shape
3. Presence of artefacts (e.g. tissue-air interface "ghosts", lipid contamination, motion artefacts, eddy currents, etc.)
4. Poor water suppression (< 98%)
5. SNR < 3 (when metabolite used for quantification)
6. SNR < 2 (for metabolite detection only)

7. CRLB > 50% (CRLB will be discussed in more detail in the section on *quantification quality control* - 5.4.2)

8. Residuals contain unexplained features (also part of *quantification quality control*)

5.4.1.2 The problems of manual quality filtering

Manual quality filtering can be performed to impose these quality criteria but comes with several problems. First, the human assessment is highly subjective and context dependent: the same signal can be classified differently depending on the quality context it is shown in (i.e. depends on the other spectra in the dataset), the level of experience of the rater, the time of the day, or even her/his mood. Second, manual assessment of the spectra requires a certain minimal degree of expertise, and, especially in clinical settings, experts are often not sufficiently available in many centres that want to apply MR spectroscopy as a clinical routine tool. Third, manual assessment of large datasets is extremely time-consuming. This can already be a problem in large collections of SVS spectra, but is considerably augmented in 2D and 3D MRSI, and is worst in high spatial resolution MRSI-techniques such as EPSI [Posse 1995, Mulkern 2001] and SPICE [Lam 2015, Ma 2016]. Consequently, very often if automatic methods for *quality control* are not available, the quality of the data is only manually checked when this is absolutely necessary, e.g. if a clinical diagnosis depends on it.

5.4.1.3 Automatic methods

More recently, automatic and semi-automatic *quality control* methods have been proposed for performing quality checks of MRS signals. An example of a semi-automated approach is presented in one of the papers [van der Graaf 2008] on the INTERPRET project. In that work, the *quality control* strategy used to assess the SVS brain tumour proton spectra of the INTERPRET database is described. The decision whether the spectra met the quality criteria for sufficient quality had two main steps. First, the values of the automatically determined linewidth and SNR were used to accept or reject spectra based on pre-defined thresholds (rejected if SNR<10 or FWHM> 8 Hz). Second, between 1 and 3 spectroscopists checked the spectra for the presence of artefacts. More details on the workflow used can be found in the original paper.

Several novel methods for automatic *quality control* can be found in more recent literature [Wright 2008, Menze 2008, Wright 2013, Pedrosa de Barros 2016b, Pedrosa de Barros 2016a]. These methods rely on pattern recognition algorithms that are able to learn to automatically assess spectral quality based on pre-labelled data that is used for training. Automatic methods have shown high performance, in some cases identical to expert spectroscopists with more than 20 years of experience. These methods present a viable solution for problems that are associated with the manual labelling: first, automatic methods are completely objective and reproducible; second, once trained, they do not require the input of a human MRS expert anymore; and third, are normally very fast, needing only a few seconds to classify the quality of hundreds of spectra. These methods may contribute to the wider application of MRS in a clinical setting, e.g. allowing the application of the techniques in clinical centres where no MR-spectroscopists are available. The reason for this is that quality filtering is one of the most important steps for quality assurance that needs to be performed obligatory on each spectrum, before the spectra can further be analysed on whether they are pathologic or not.

Wright *et al.* [Wright 2008] presented in 2008 an automatic method for assessing MRS quality in single voxel brain tumour spectra. The method was trained and tested on SVS data from the eTumour [Julià-Sapé 2012] and the INTERPRET databases that had been previously labelled as either "acceptable" or "non-acceptable". The strategy used for the assessment was the previously mentioned semi-automatic approach used in the INTERPRET project [van der Graaf 2008]. The method followed a 2-step approach where in a first step an Independent Component Analysis was used to reduce the dimensionality of the problem, and in a second step the resulting features were used in a least-squares support vector machine. The independent components used were fixed prior to the testing of the method. The method showed a performance similar to the experts, given that it had an average agreement of 87% with the experts' judgements, and that the average agreement across all pairs of experts was 85%.

The same method was later on successfully applied to MRSI data of prostate tumours [Wright 2013]. Also in this study, the agreement of the classifier with the experts was similar to the agreement between experts. Besides the excellent classification results, the data selection was focused on consensus labelling and, only those cases for which experts reached agreement upon, were used to train and test the classifier. Experts agree almost every time on the "clearly good" and "clearly bad" cases but disagree often in those cases that fall in the boundary between acceptation and rejection. This might make it tempting to avoid these cases and interpret them as "sources of inaccuracies" for the classifier. Learning theory shows the opposite: the cases with most uncertainty in class membership carry the most information for the classifier [Lewis 1994].

Another automatic classifier for MRSI data was presented by Menze *et al.* [Menze 2008] in 2008. This approach used the Random Forest (RF) [Breiman 2001, Liaw 2002] classifier and tested the approach on long-TE PRESS data of brain tumour patients. The RF-classifiers used as input features the magnitude-spectra intensities in the range between 0.5 and 3.6 ppm. The performance obtained for the spectra of the excitation volume had an area under the curve (AUC) between 0.950 and 0.968, depending on whether borderline-quality spectra were considered *acceptable* or not. The classifier was compared against SNR and Cramér-Rao-Lower-bounds thresholding, showing consistently a higher performance.

More recently, our group presented a method [Pedrosa de Barros 2016b, Pedrosa de Barros 2016a] for automatic control of MRSI brain tumour spectra that relies also on the RF-classifier but uses a different set of features. Instead of using directly the points of a certain chemical shift range of the spectra, a total of 47 different signal features were extracted from both time and frequency domain. This work not only allowed to obtain very accurate automatic classifiers (AUC in long TE 0.9769 and short TE 0.9624) that are as accurate as an expert, but also to identify several signal features that are highly correlated with spectral quality.

5.4.1.4 Signal features with relevant information for automatic quality assessment

Figure 5.7 shows the feature importance of the several features used in the proposed method. The measure used to evaluate the feature importance was the error increase seen in the performance of the classifier after permuting the values of a selected feature between the different examples of the test set. A higher increase in the error after permuting the values of a certain feature is interpreted as higher importance for the classification (more information about this measure can be found in [Breiman 2001, Liaw 2002]). Seven features are worth mentioning in more detail due to their high impact in the classification accuracy. First the SNR of the

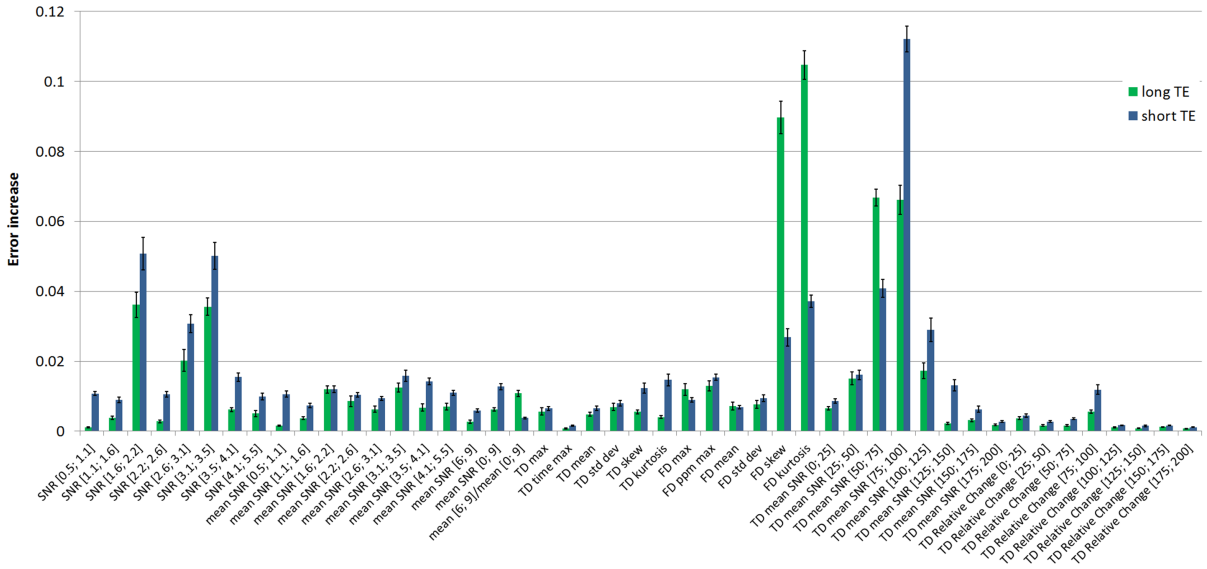

Figure 5.7: Feature importance plots. The vertical axis shows the mean error increase observed after permuting the values of the corresponding feature in the dataset [Liaw 2002]. The mean and standard deviation of the error increase were calculated from the 12 different random forests that were trained during cross-validation. The error bars show the mean ± standard deviation of the error increase. FD, frequency domain; SNR, signal-to-noise ratio; TD, time domain. For more details on the features and cross-validation strategy see [Pedrosa de Barros 2016b].

highest peak found in the ranges between 1.6-2.2 ppm, 2.6-3.1 ppm and 3.1-3.5 ppm. For almost every brain spectrum, these correspond respectively to the SNR of NAA, creatine and choline. As expected, these features correlate well with signal quality. However, these features were not identified to be the most important features by the computed RF-classifiers. In frequency-domain, the skewness and the kurtosis showed higher importance for long TE data. In time-domain, the mean SNR in the ranges between 75-100 ms for long-TE, and 50-75ms and 75-100ms for short-TE, was also more important than the metabolite SNR.

5.4.1.5 Frequency-domain skewness and kurtosis

Skewness and kurtosis are measures of, respectively, the third and fourth moments about the mean. According to Cramér [Cramér 1946, Joanes 1998] these can be estimated for a sample of n values using the following expressions:

$$skewness = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^3}{\left(\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2\right)^{\frac{3}{2}}} \tag{5.1}$$

$$kurtosis = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^4}{\left(\frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2\right)^{2}} - 3 \tag{5.2}$$

Skewness measures the symmetry of a distribution, being zero for symmetric distribution and increasing its magnitude with increasing asymmetry. Skewness can be positive or negative depending on whether the distribution is left or right skewed. On the other hand, kurtosis reflects the "peakedness" of the distribution, or more correctly, the weight of its tails. Heavy-tailed distributions show negative kurtosis, whereas light-tailed distributions have positive kurtosis. The formula presented here for kurtosis uses as reference the normal distribution,

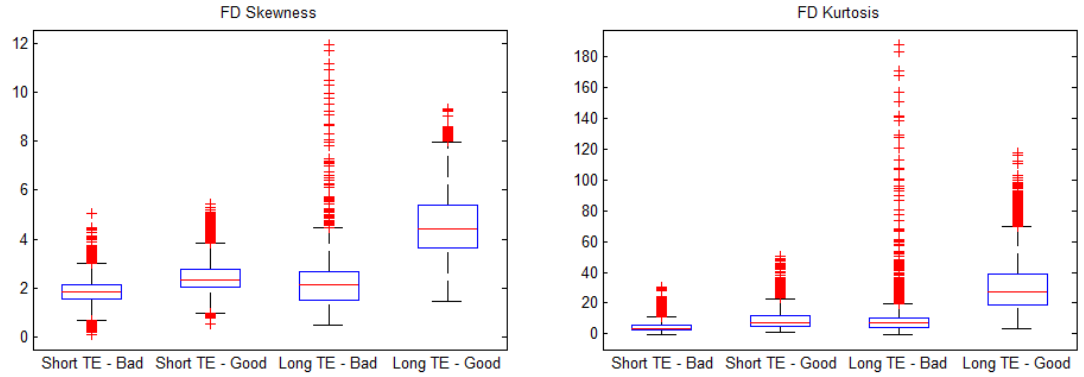

Figure 5.8: Box and whiskers-plot showing the values of the FD skewness (Left) and FD kurtosis (Right), grouped by quality label ("Good" or "Bad") and TE (short-30ms and long-135ms). The features were extracted after residual-water-peak removal. A total of 19032 MRSI spectra are represented in these plots (3708 Short TE "Bad", 5808 Short TE "Good", 3214 Long TE "Bad", and 6302 Long TE "Good").

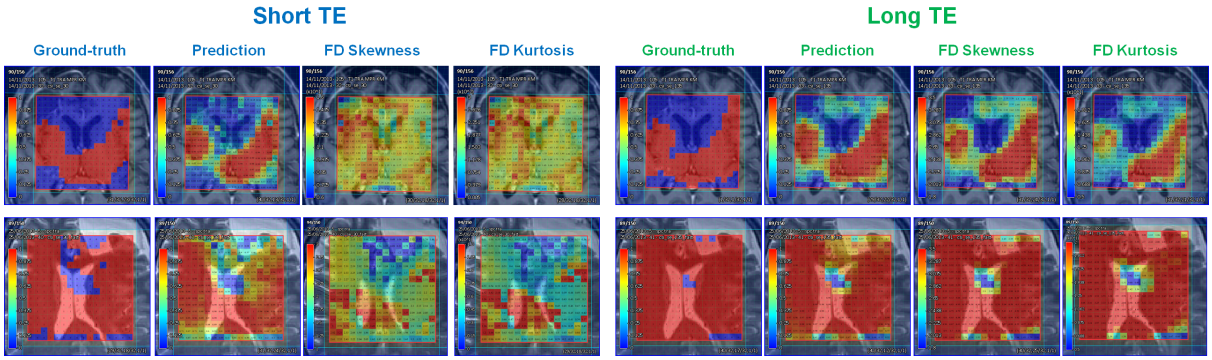

Figure 5.9: Comparison between the ground-truth (manual-labelling performed by the experts: 0 = "reject", 1 = "accept"), the prediction of the automatic classifier (raters' acceptance probability), and frequency-domain skewness and kurtosis, for both short (30ms) and long-TE (135ms) data. Each row corresponds to a different patient. Short and long-TE data were acquired sequentially and in the same exact location. Acquisition details: 1.5T Siemens Aera 2D-PRESS TR 1500ms, 12x12 k-space samples, interpolated resolution 32x32.

whose ratio between the 4^{th} and the squared 2^{nd} moment about the mean is 3. Therefore, using this formula, samples with a close to normal distribution will have a kurtosis close to zero.

Skewness and kurtosis of a spectrum can be used to assess the distribution of its intensity values. Spectra without a strong baseline and with narrow linewidths have a higher number of low level intensities and consequently have high skewness and kurtosis levels. On the other hand, baseline distortions, broad linewidths and broad signal artifacts lead to flatter histograms and, consequently, to lower skewness and kurtosis. These relations help to understand why these two features are good quality indicators for long TE data, namely after residual water peak removal, given that the residual water causes considerable histogram changes that are often not related with signal quality. In the case of short TE, skewness and kurtosis are highly reduced by the strong macromolecular baseline independently of the quality of the spectra, and consequently these features contain less information for the identification of poor-quality spectra. Figures 5.8 and 5.9 show the relation between these features and the quality of the spectra for both short and long TE.

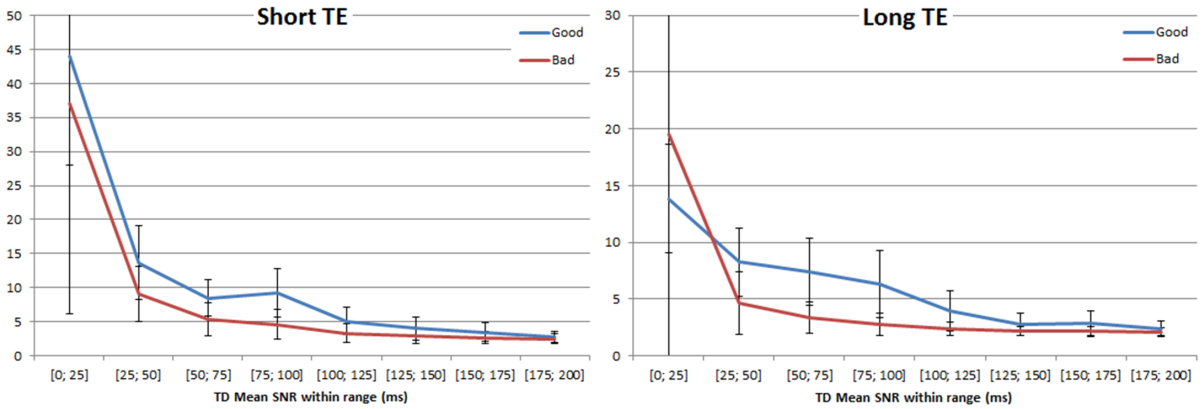

Figure 5.10: Plots of the Time-Domain Mean SNR for different ranges, for both short and long TE data. The plots show the average value ± standard deviation over all "Good" or "Bad" spectra of the dataset.

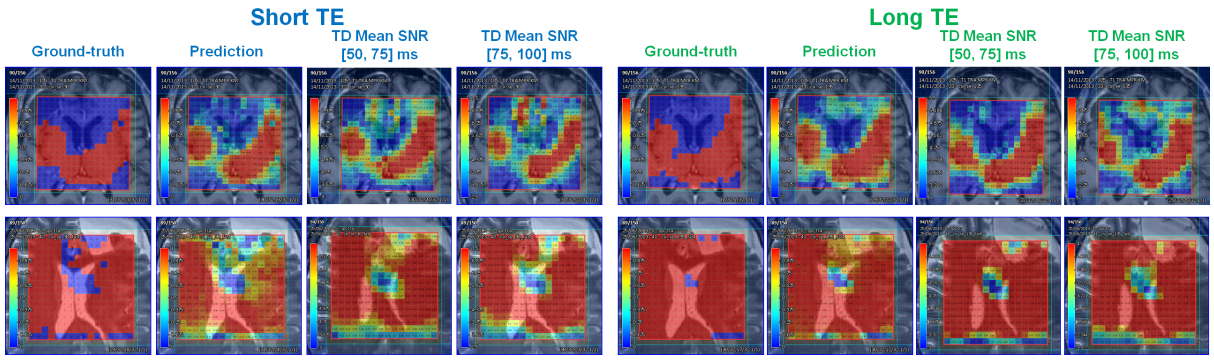

Figure 5.11: Comparison between the ground-truth (manual-labelling performed by the experts: 0 = "reject", 1 = "accept"), the prediction of the automatic classifier (raters' acceptance probability), and time-domain mean SNR within the 50-75ms and the 75-100ms ranges, for both short (30ms) and long-TE (135ms) data. Each row corresponds to a different patient. Short and long-TE data were acquired sequentially and in the same exact location. Acquisition details: 1.5T Siemens Aera 2D-PRESS TR 1500ms, 12x12 k-space samples, interpolated resolution 32x32.

5.4.1.6 Time-domain mean SNR for different ranges

Time-domain mean SNR is another interesting feature, namely in the range between 50 and 100 ms for the tested data (Siemens Aera and Avanto, 1.5T, PRESS, TR 1500ms, TE 135 and 30 ms). The relation of this parameter with signal quality is related with the shorter $T_2{}^*$ values of the signals with poor quality. Poor quality signals, even those with extremely high amplitude artefacts, decay more rapidly than others with higher quality. Consequently, the most relevant time domain range to evaluate signal quality is the range where poor quality FIDs have already decayed almost completely but where FIDs with longer $T_2{}^*$ have still moderate intensities (Figure 5.10 and 5.11). This range should be dependent on the acquisition conditions, namely field strength. In contrast with skewness and kurtosis, time-domain mean SNR is an important quality indicator for both short and long TE data.

5.4.2 Quantification quality control

Several metrics have been used to assess the reliability of the quantification results. In this section we will discuss some of those measures, namely the Cramér-Rao Lower Bounds (CRLB) [Jiru 2006, Ratiney 2005, Cavassila 2001], methods for experimental error estimation such as the ER-ARSOS [Slotboom 2009, Slotboom 2011], and the fit quality [Slotboom 2009, Slotboom 2011]. These measures can be used to detect the situations where the quantification method used, due to several factors such as low SNR or line-shape distortions, was not able to generate accurate results.

5.4.2.1 Cramér-Rao Lower Bounds

Cramér-Rao Lower Bounds (CRLB) [Jiru 2006, Ratiney 2005, Cavassila 2001] provide a *theoretical* estimation of the minimum obtainable variance in the estimation of a model-parameter, given the measurement data. For a given set of model parameters

$$\theta = [\theta_1, \theta_2, \ldots, \theta_n]^T \tag{5.3}$$

the Cramér-Rao inequality for the parameter i, considering an unbiased estimator $\widehat{\theta}$ is given by

$$var\left(\widehat{\theta_i}\right) \geq \left[I^{-1}(\theta)\right]_{ii} \tag{5.4}$$

where $\left[I^{-1}(\theta)\right]_{ii}$ is the i–th element of the diagonal of the inverted Fisher information matrix, and $var\left(\widehat{\theta_i}\right)$ the variance of the estimator of the parameter i.

One obtains high CRLB values, for example, in cases of low metabolite SNR, or (strongly) overlapping resonances of two or more resonance lines. CRLB error estimation is only valid if the assumption that the model used to describe the data fully applies to the data, and the correct probability density function of the noise is known. Therefore, in case signal artefacts are present or the spectra contain more resonances than modelled, the error estimation provided by the CRLB is *not* valid [Slotboom 2011]. Another important property of the CRLB is that the error estimation increases with the number of estimated parameters. For this reason, the number of parameters to estimate should be kept as low as possible. The most important way to keep the number of parameters as low as possible is to make use of *prior knowledge* models.

Many groups use quality filtering based on relative CRLB (ratio between the CRLB and the estimation of the parameter). Despite its common use, the use of relative CRLB instead of absolute CRLB comes with several dangers. As explained in the work of Kreis [Kreis 2015],

the usage of relative CRLB filtering may strongly bias study data, since metabolites having low concentration values will be classified as poor quality data. Moreover, the analysis of data from conditions leading to low metabolite concentrations might become completely impossible, since most of the data from the patients with such condition would be potentially rejected.

CRLB are available in quantification methods such as LCModel [Provencher 1993] and QUEST [Ratiney 2005] and several examples of their use can be found in literature. An example of the use of CRLB is given in the work of Jiru *et al.* [Jiru 2006], where the authors suggest that metabolite maps should always be accompanied by a CRLB map to help identifying spectra with lower quality.

5.4.2.2 Experimental Error

For SVS, where signal averaging is applied to obtain sufficient SNR, the quantification error can be estimated experimentally [Slotboom 2011, Bolan 2004, Bolliger 2013], as an alternative to the CRLB. To determine the error, quantification is performed independently for the several signals, and the standard deviation in the quantification of each metabolite is then calculated. Since the number of signals might be, in some cases, too small for obtaining accurate estimates, Bolliger et al. [Bolliger 2013] proposed the use of *bootstrapping* to increase the size of the dataset artificially. The author showed that errors estimated from a *bootstrapped* sample are very close to the theoretical estimations provided by CRLB.

ER-ARSOS [Slotboom 2009, Slotboom 2011] is another method that uses separately stored signal acquisitions for experimental error estimation. In this technique, the error of the measurement is estimated by the variance in the quantification results of the separately filtered signals. The filtering is performed by means of All Rank Selected Order Statistic filtering (ARSOS) [Slotboom 2009] prior to the quantification. This unique filter turns the N low-SNR individual input signals, into $N/2$ high SNR output signals, which can all be quantified meaningfully. For *in vivo* MRS, the observed ER-ARSOS errors that are found, are mostly larger than the CRLB which are often too optimistic. The reason for this is that ER-ARSOS not only depends on Gaussian noise, as CRLB does, but also on heavily tailed noise that scales with the MRS signal itself. Examples of sources of heavily tailed noise include, for example, patient motion and tissue pulsation.

5.4.2.3 Fit Quality Number

Another practical measure of the error associated with the quantification is given by the "normalized fit quality number" [Slotboom 2009, Slotboom 1998]:

$$Q_{fit}(N) = \frac{R}{N \cdot \sigma} \tag{5.5}$$

, where N is the number of data points from the time or frequency domain signal used, R is the sum of the squares of the residuals (difference between the fit and the measured signal), and σ the variance of the signal measured in a region with no metabolite signal contributions. This measure is close to 1 for a close to perfect fit. Values of Q_{fit} significantly larger than 1 indicate significant differences between the fit and the experimental data which could indicate an inappropriate data model or a local minimum in the sum of least squares (χ^2). On the other hand, Q_{fit} values smaller than 1 strongly indicate *overfitting* of the model to the data.

5.5 Techniques to Improve Quality

Quality of MR spectra is a very broad subject, and many different strategies can be found in the literature for tackling different quality issues. In this section, we mention some techniques and corrections that can be used to improve the quality of the spectra.

5.5.1 Eddy current correction

Eddy currents cause the frequencies of all resonances to become time and spatial dependent, leading to complex distortions of the line-shape (so called "wiggles"). If not compensated, these distortions give rise to quantification errors. From the scanner hardware side, shielded gradient systems, in combination with active electronic eddy current compensation systems are used to minimize their effects [Van Vaals 1990]. Even after taking these measures to minimize their effects, often residual eddy currents effects are still visible in the spectrum. In order to correct these undesired residual eddy current effects, de Graaf *et al.* proposed already in 1990 a deconvolution method named QUALITY (**qua**ntification improvement by converting **li**ne-shapes to the lorentzian **ty**pe) [De Graaf 1990]. This method performs on the acquired MR-signal a time domain deconvolution with a reference signal, normally the unsuppressed water signal of the same localization, allowing to obtain perfect Lorentzian line-shapes.

5.5.2 Elimination of unwanted MRS Signals: Phase Cycling and Crusher Gradients

The flip-angles of selective RF-pulses, which are used in localized MRS, are non-ideal, i.e. these pulses do *not* excite the spins uniformly over the excited target volume. Besides this, the pulse imperfections lead also to signal excitation outside the target volume, namely for flip angles larger than 90 degrees. Direct application of these pulses, without taking any actions that eliminate the effects of these imperfect flip angles, results in the presence of unwanted echoes that cause strong artefacts in the localized spectra. In order to prevent these artefact signals, two techniques are normally used: RF-pulse phase cycling (e.g. EXORCYCLING, [Bodenhausen 1977]), and the use of "crusher gradients pairs" that actively dephase all unwanted MRS-signal contributions that were excited by the RF-pulse. These two techniques are normally combined to eliminate the unwanted echoes even more effectively.

5.5.3 Order of selection pulses

A very simple strategy that can lead to a drastic quality improvement in sequences such as PRESS and STEAM, is the correct decision over the order of the selection pulses. This topic is extensively discussed in the work of Ernst T. and Chang L. from 1996 on the elimination of signal artifacts in short TE brain spectra of the frontal lobe [Ernst 1996].

"Ghost" artefacts are often seen near the sinuses, namely in the frontal lobe. These correspond to signals from insufficiently dephased unwanted coherences originating from the sinuses and the mouth. As explained in the mentioned reference, the origin of these unwanted coherences depends on the exact order in which the three slice-selective pulses are applied. Therefore, given that the third slice selection pulse is more subjective to insufficient spoiling, this should cover the "least problematic" plane, and avoid regions prone to field inhomogeneities such the as the sinuses and the mouth. Consequently, in order to improve the quality of the spectra acquired in the frontal lobe, the axial plane should be selected at last. Figure 5.12

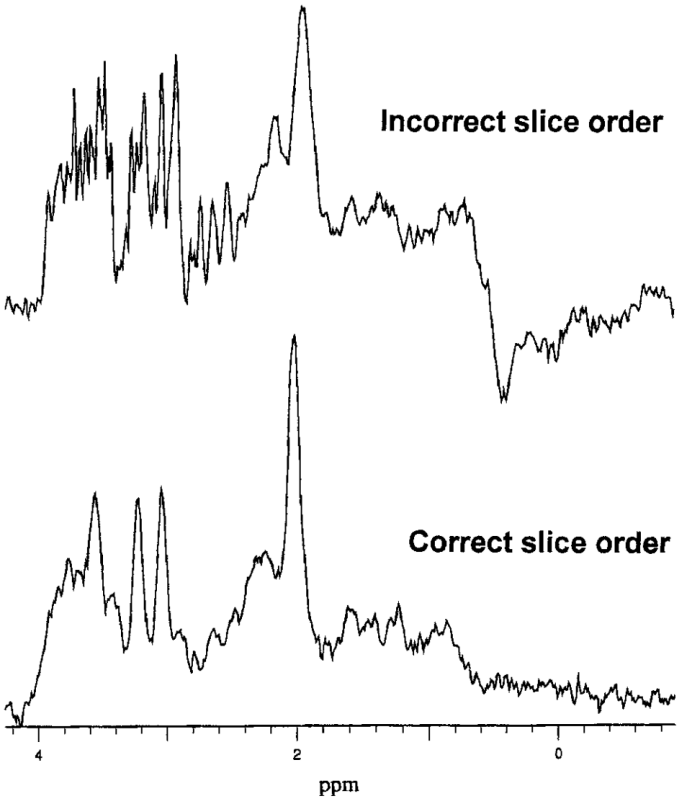

Figure 5.12: SVS spectra from the midfrontal lobe of a healthy volunteer. The slice order of the upper spectrum was axial-coronal-sagittal, whereas from the lower was coronal-sagittal-axial. Original figure from [Ernst 1996]. Copyright by John Wiley and Sons.

shows a good example of the impact that the slice order can have on the quality of the obtained localized spectra.

5.5.4 Chemical Shift Displacement Artefact

Given that the spins of the different substances measured in MR spectroscopy have different chemical shifts, and consequently different resonance frequencies, the region that is selected for each spin by the same RF pulse depends on its chemical shift. The resulting displacement in the localization of the VOI caused by the differences in the chemical shift is referred to as the Chemical Shift Displacement Artefact (CSDA).

The relation between the resonance frequency and the corresponding generic coordinate r, depending on the gradient in the r-direction G_r, the chemical shift σ, and the main magnetic field B_0 is given by the following expression [Goelman 2007]:

$$\omega(r, G_r, \sigma) = \gamma G_r r + \gamma B_0 (1-\sigma) \Leftrightarrow r(\omega, G_r, \sigma) = \frac{\omega - \gamma B_0 (1-\sigma)}{\gamma G_r} \tag{5.6}$$

, where γ is the gyromagnetic ratio. From this expression, it follows that the displacement Δr caused by the chemical shift, for a fixed frequency ω'_{RF} of the RF-pulse, is given by:

$$\Delta r = \frac{\omega'_{RF} - \gamma B_0(1-\sigma_2) - \omega'_{RF} + \gamma B_0(1-\sigma_1)}{\gamma G_r} = \frac{\gamma B_0 \Delta\sigma}{\gamma G_r} = \frac{\gamma B_0 \Delta\sigma}{BW_{RF}} FOV_r \tag{5.7}$$

, where BW_{RF} is the bandwidth of the RF-pulse, and FOV_r the Field-of-View in the r-direction. This expression shows that the chemical shift displacement increases linearly with the magnetic field strength and the FOV, and decreases with the bandwidth of the RF-pulse (see Figure 5.13). Therefore, the use of higher bandwidth pulses is an effective way of reducing the displacement effect. However, it should be noted that for non-adiabatic, i.e. amplitude modulated only, RF-pulses, the excitation bandwidth is limited by the RF-amplifiers maximum available RF-peak power. At the same RF-peak power, adiabatic pulses can obtain much larger excitation bandwidths than non-adiabatic RF-pulses.

Figure 5.14 shows the effect of the CSDA for a 2D-MRSI PRESS sequence, where the selection profile of choline is shifted in the y-direction relatively to that of lactate. Knowing that the bandwidth of the 90° pulse is considerably larger than the one of the 180° pulses, it can be concluded that the 90° pulse was applied in the x-direction.

The choice of the orientation of the different pulses should consider the impact that the chemical shift can have in each direction. For example, in 2D-MRSI slices of the brain acquired using PRESS, the use of the 90° pulse in the x-direction (from right to left of patient) reduces the impact that chemical shift has on the contralateral comparison. Moreover, for the same example, given that the FOV in the z-direction is considerably smaller than in the other directions, it is advisable to use one of the 180° pulses in that direction.

The CSDA can lead to signal cancellation in coupled spin systems [Slotboom 1994b, Lange 2006, Kaiser 2007] when pulse echo sequences are used. This results from the fact that the differences in displacement between the coupled spins of the same molecule (e.g. for lactate) will create different compartments where, for each direction, both or only one of the spins is selected (Figure 5.15). The evolution under J-coupling of these non-uniformly excited spin systems will be different in each physical compartment, which may lead to the creation of MR-signals with opposing phase, causing intra-voxel signal annihilation. This effect can lead to a considerable underestimation of certain metabolites, such as lactate [Lange 2006]

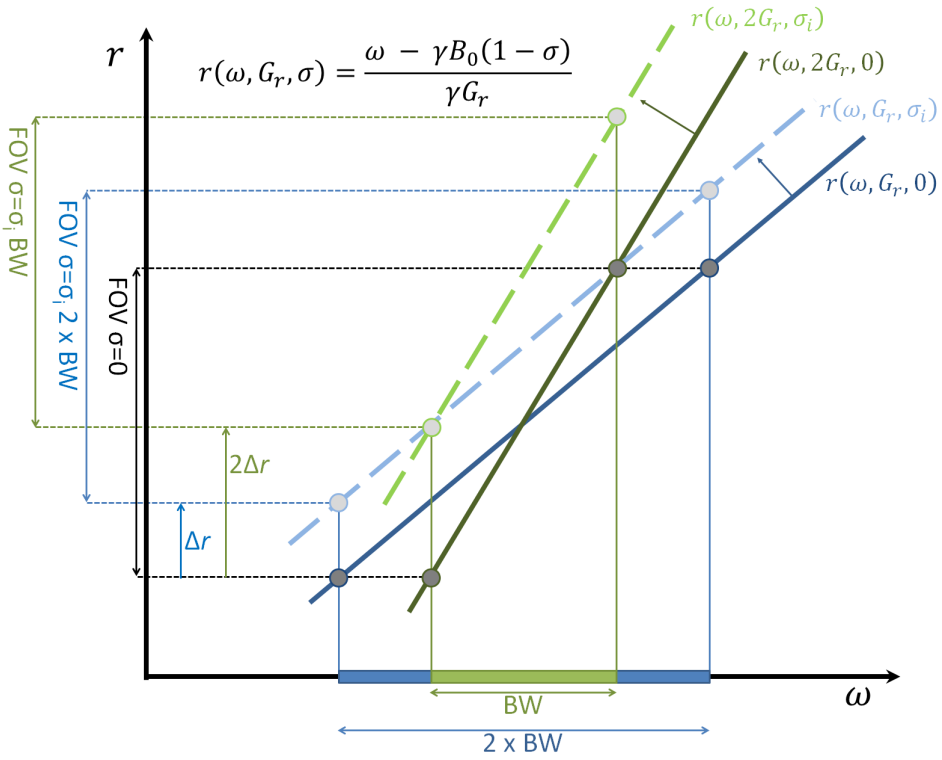

Figure 5.13: Plot showing the effect of doubling the bandwidth of the RF-pulse on the chemical shift displacement, for a given chemical shift difference. The solid lines show the relation between frequency and the generic coordinate r for the protons resonating at the Larmor frequency. In these lines, the gradient is adapted so that the same region is selected independently of the bandwidth of the pulse. The 2 dashed lines show how the relation between frequency and the selected coordinate changes for a chemical shift $\sigma_i > 0$. The selected FOV depending on the RF-pulse bandwidth and chemical shift is displayed on the vertical axis.

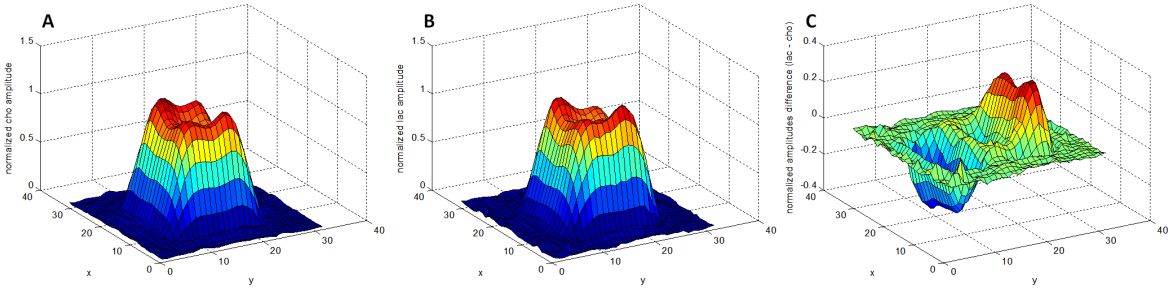

Figure 5.14: 2D selection profiles of Cho (A) and Lac (B) and difference between them (C) showing chemical shift displacement in the y direction for a 2D PRESS sequence (1.5T Siemens Aera, TE/TR 135/1500ms). The amplitudes of each metabolite were normalized so that the selection profiles have the same integral as a perfect step function selecting the same VOI (amplitude 1 inside VOI and 0 outside).

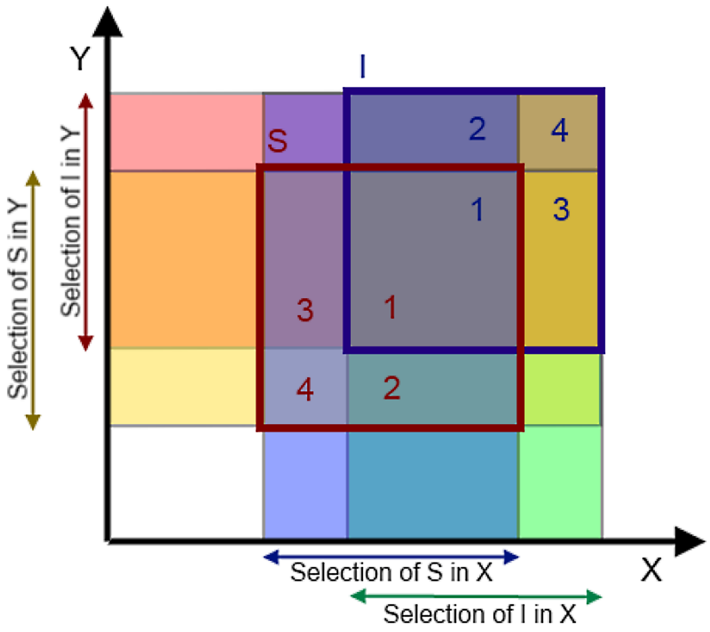

Figure 5.15: Different localized regions in the presence of chemical shift displacement for two spins I and S with different chemical shift. The image shows the different selection compartments in the xy-plane when localization is applied in the x and y directions. The four distinct localized regions from each spin can be distinguished in the figure.

and GABA [Kaiser 2007]. Sequences such as SADLOVE or semi-LASER offer an effective solution for this artefact [Slotboom 1994b], given that the phase-compensated 2π pulse pair [Conolly 1989b, Conolly 1989a] have a considerably large bandwidth, and the effects of J-coupling during the 2π pulses are minimal (see Figure 5.16).

5.5.5 Metabolite cycling

Another technique that improves the quality of the spectra in challenging areas like the spinal cord, is the so-called "*metabolite cycling*". The use of this technique and the impact that it can have on the quality of the recorded data is well exemplified by the work of Andreas Hock and his colleagues on MR-spectroscopy of the spinal cord [Hock 2013]. Metabolite cycling consists in the interleaved acquisition of spectra with and without the use of an inversion pulse that targets the metabolite signals. Consequently, it is possible to separate water and metabolite signals without performing water-suppression. Moreover, since the water signals are not distorted by the water-suppression RF-pulses, they can be used for eddy-current correction, as well as to correct the frequency shift of the metabolite signals caused by the patient motion and B_0 field drift. The authors demonstrated the effect of performing these corrections before signal averaging in the final result. The article also nicely demonstrates a comparison between spectra acquired with metabolite cycling, and spectra acquired with water-suppression, showing that the first has significant narrower linewidths and higher SNR.

5.5.6 Strategies for tackling movement artefacts

An important source of error in any MR acquisition is patient movement. As shown in the work of Kreis [Kreis 2004], movement artefacts are not easily detected in SVS after signal averaging, not even when the patient leaves the scanner in the middle of the acquisition. Several authors proposed different strategies to tackle movement artefacts.

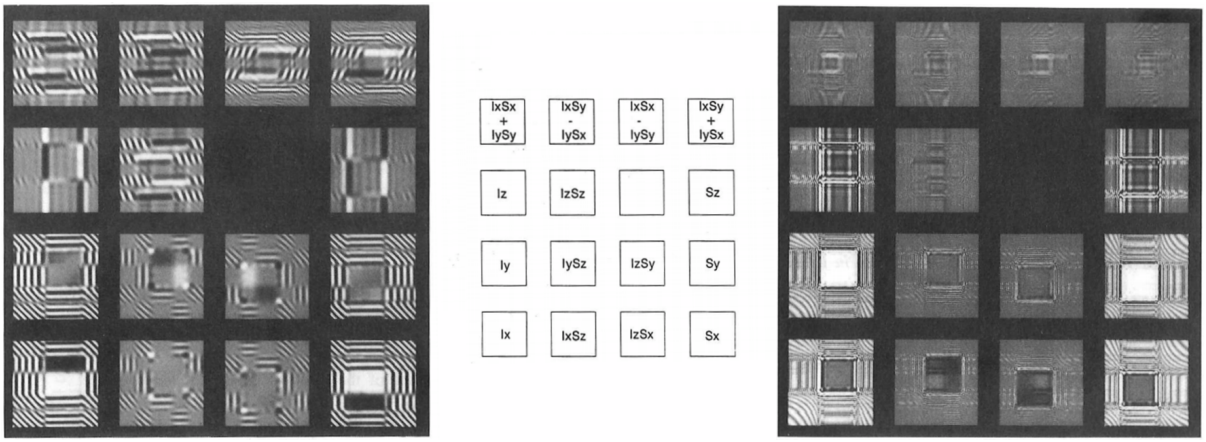

Figure 5.16: Spatially dependent coherences for 2 spin coupled system with coupling constant $J = 7Hz$ resulting from the PRESS localization (left), and SADLOVE localization (right). The identification of the corresponding product operators is presented in the middle. In the image, the grey level corresponds to the expectation value of the phase: white represents 1 and black -1. It was considered a bandwidth of 2kHz for the selective pulses and total offset range of 4.5 kHz in both x and y directions. Especially note that in the PRESS experiment there is a 180° phase flip over the selected volume, which results in signal annihilation. The SADLOVE (later LASER) pulse sequence does not show this effect for the same spin system. In order to make the results comparable, the SADLOVE sequence used was a variant of the original sequence where the first pulse was replaced by a hard 90° pulse (similar to semi-LASER). Source: [Slotboom 1993].

Slotboom *et al.* introduced the ARSOS filtering technique described above [Slotboom 2009] to detect signal outliers in the separately stored signals acquired during the process of signal averaging. ARSOS orders the N-equivalent time point or frequency point values of the separately stored acquired signal repeats point-by-point for all points of the TD or FD signals. After this reordering, the output signals close to the extremes (minimum and maximum output) contain the outlier signal contributions, often dominated by the effects of movement and tissue pulsation. The filter works *similarly* to a median filter, and presents a straightforward solution to tackle movement without having to perform any changes to the sequence. Depending on the number of outlier signals containing filter outputs, ARSOS normally reaches higher SNR than application of conventional median filtering. The only requirement for using ARSOS is that the different signals are stored separately.

Another solution to detect movement artefacts is proposed by Bhattacharyya *et al.* [Bhattacharyya 2007] who suggest the use of an interleaved acquisition of water spectra to detect movements. The method was applied to a MEGA-PRESS sequence, where movements can lead to undesired differences between ON and OFF acquisitions. Besides allowing the detection of spectra affected by patient movements, the non-suppressed water spectra can be used to perform eddy current and frequency shift correction, similarly to what was done in the metabolite cycling method described above.

A more advanced approach to tackle movement artefacts is the vNav EPI suggested by Hess *et al.* [Hess 2011, Hess 2013]. The method consists in the interleaved acquisition of an EPI navigator, which is used to correct, in real time, the position of the spectroscopic voxel. This correction of the localization is then used to perform frequency adjustment as well as first-order shimming. The navigator is acquired during the recovery time, and has no impact in the total acquisition time. The examples presented by the authors show the clear added robustness that this approach can provide to a PRESS sequence.

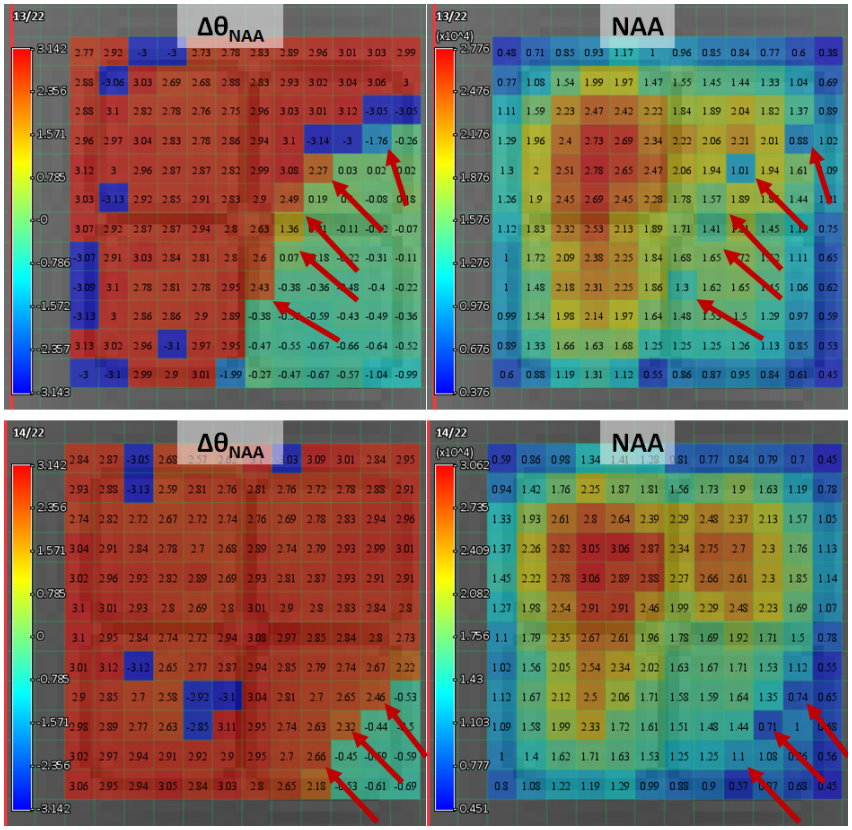

Figure 5.17: Example showing signal cancellation due to incorrect phasing of the metabolite signals prior to coil combination. Between the acquisitions of the first and the second row the shimming parameters were updated causing the region, where a transition in the phase of NAA is observed, to move its location. The arrows signal the transition region. $\Delta\theta_{NAA}$ corresponds to the phase difference between the original signals (phased using first point of FID and without removing residual-water-peak) and the same signals after removing the residual-water-peak and correctly phasing NAA. NAA maps calculated by peak integration of magnitude signals.

5.5.7 Metabolite signal cancellation due to incorrect phasing of water-suppressed signals

Water suppression sequences such as CHESS[Haase 1985], consist of pairs of frequency-selective pulses and dephasing gradients. These sequences are optimized such that, for the average T_1 of water, the longitudinal magnetization (M_Z) is nulled by the time the first excitation pulse of the volume-selection takes place. Given that M_z is never completely nulled, and that the concentration of water is several orders of magnitude greater than other substances, the amplitude of the remaining water signal is often greater than the amplitude of the other signals. Moreover, in case M_z is negative at the start of volume selection, water will be acquired with an inverted phase relatively to the metabolite signals. The phase of the water signal depends on the effect that deviations in both T_1 of water and flip-angle of the pulses have on the water-suppression sequence used. Since these phase deviations jeopardize the use of the water peak for frequency-shift and phase correction, the time between the water saturation and start of the volume selection can be slightly extended such that the water spins have more time to relax towards equilibrium, resulting in a situation where the residual water and metabolite resonances have the same phase. On Siemens scanners, this type of water suppression is referred to as “weak water suppression”.

Due to above mentioned effects, in MRSI water-suppression may produce regions where

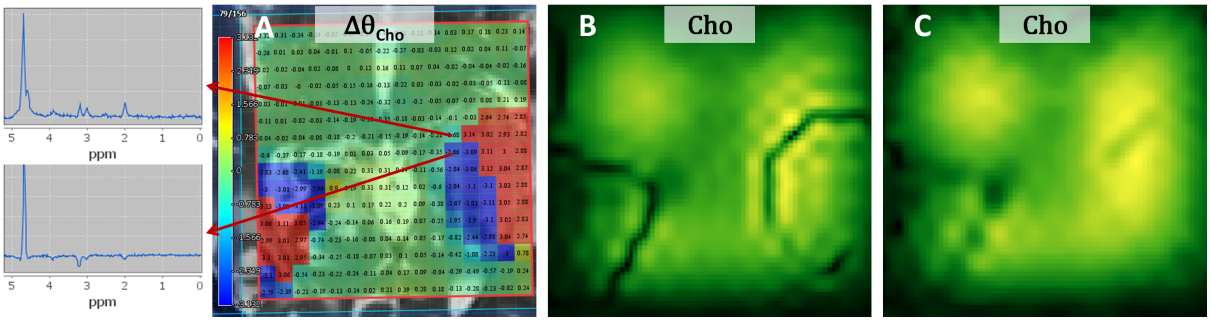

Figure 5.18: A) Phase map of choline for a 32x32 dataset. $\Delta\theta_{Cho}$ calculated as the phase difference between the original signals (phased using first point of FID and without removing residual-water-peak) and the same signals after removing the residual-water-peak and correctly phasing Cho. The spectra of two adjacent voxels are presented: in the first, water and metabolites have the same phase whereas in the second, the phase of the water and the metabolite signals differ. B) Interpolated (128x128) choline map showing signal cancellation due to phase jumps between adjacent voxels. C) Interpolated (128x128) choline map after residual-water-peak removal and auto-phasing. The correction in this case is not complete given that the MRSI grid had already voxels with signal cancellation due to coil-combination of badly phased metabolite-signals.

the phase of the water (relatively to the metabolite signals) abruptly changes sign from one voxel to another. In those regions, metabolite-signal cancellation is frequently observed. This is seen in Figure 5.17, where two phase maps and corresponding NAA maps are shown. These two examples were acquired using the same phantom and acquisition parameters but with different shimming parameters. The update in shimming caused the phase-jump region to shift its position, and consequently the signal cancellation moves with it.

The signal cancellation observed in Figure 5.17 can be explained by the combination of incorrectly phased metabolite signals from different coils. Incorrect phasing is predominantly caused by differences between the phase of the residual water signal and the metabolite signals, and is especially pronounced when zero-order phasing is performed globally, using for example the first point of the FID. The result of auto-phasing is very sensitive to signal differences in the regions where phase-jumps are observed. Consequently, signal-combination in phased array coils performed after incorrect phasing leads to signal cancellation in those regions [Strasser 2012].

Besides the above mentioned mechanism of signal cancellation, cancellation can also occur as a result of spatial interpolation by means of zero-filling, as depicted in Figure 5.18-B. In this case, signal cancellation occurs between neighbouring voxels that have metabolite signals with opposing phase. Figure 5.18-C shows the interpolation of the same MRSI grid when the data is previously phased after removing the residual water using HLSVD, therefore preventing metabolite-signal cancelation due to interpolation. Even better, if interpolation is performed on the quantification maps, since phase is handled separately for each metabolite by the fitting algorithm, signal cancellation can be prevented.

In summary, signal cancellation can be prevented mainly by two measures. Firstly, the use of water-suppression techniques that do not create spatially dependent phase-jumps in residual-water signals. Secondly, the use of auto-phasing methods prior to coil-combination that allow for a robust phasing of the metabolite signals, regardless of the phase of the residual-water-signal.

5.5.8 Correction of slice selection profile in MRSI data

A big practical problem for the unambiguous interpretation of MRSI-data is the non-homogeneous excitation of the spins within the excited volume of interest. These variations

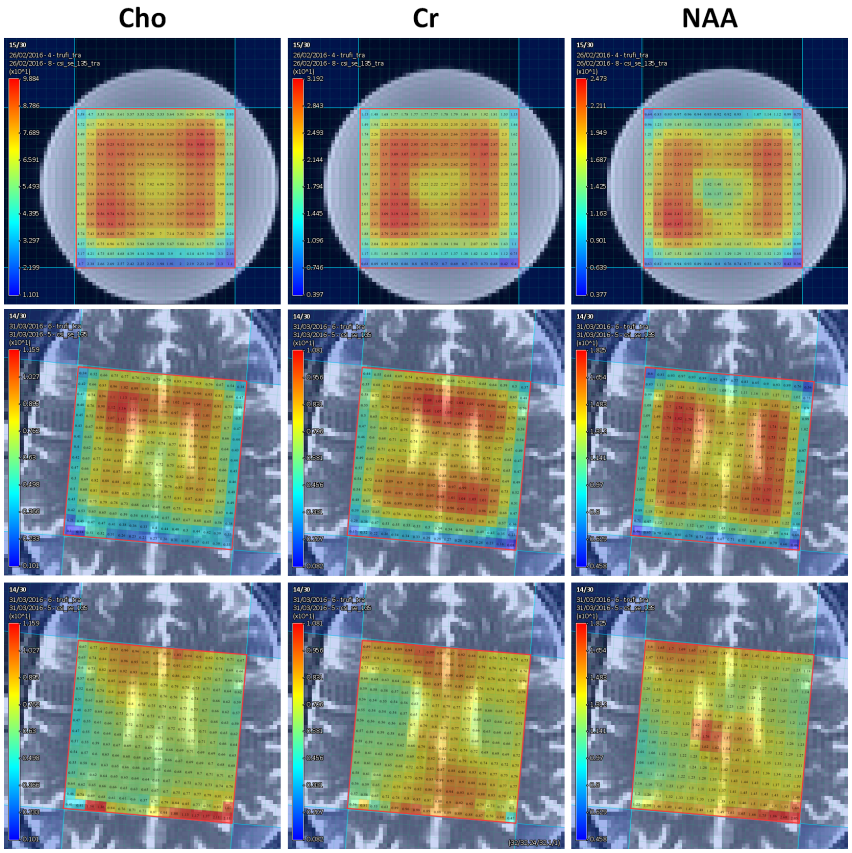

Figure 5.19: Slice profile correction. The top row shows the metabolite maps acquired in a phantom containing a solution of 17.5 mM Cho, 17.5 mM Cr and 17.5 mM NAA, and shows the amplitude of the three metabolites measured using QUEST. In the middle row, we see three metabolite maps (quantification using QUEST) of a healthy test person. The last row shows the *in vivo* metabolite maps after correction using the normalized maps of the first row.

Table 5.2: Coefficient of variance for choline, creatine and NAA, before (second row Figure 5.19) and after correction (third row Figure 5.19) of the slice profile.

	Cho	Cr	NAA
Before	30.32%	28.59%	23.82%
After	24.59%	14.90%	12.76%

are clearly visible in the metabolite maps of the top row of Figure 5.19, from data acquired in phantoms with homogenous metabolite in water solutions. The variations in observed MRSI signals over the voxels are dominated by the excitation profile of the RF-pulses. Under the assumption of equal average pulse flip angles, given that the slice profiles are constant for the same RF-pulses, the spatial signal amplitude distributions of these phantom data can be used for correction of *in vivo* MRSI-data acquired using the same sequence and parameters. Since the pulse profile depends on the exact chemical of the spin, the correction factors must be determined for each resonance line individually. The two last rows of Figure 5.19 show three metabolite maps from a healthy volunteer, *before* and *after* correction. The coefficient of variance was measured for each *in vivo* map, showing that the correction of the slice profile leads to a reduction in the spatial variations of the map (Table 5.2). Further developments are in progress to incorporate the correction of these effects into a processing pipeline of clinical routine MRS.

5.6 Conclusion

This article covered several important aspects of quality management in localized *in vivo* MR-spectroscopy. First the main principles behind *quality planning*, *assurance*, *control* and *improvement* were discussed. After this, several phantom tests were described in section 5.3, and in section 5.4, methods for detecting errors in the acquisition and quantification were presented. Finally, several practical suggestions for *quality improvement* were given.

To summarize, achieving good results in MRS does not depend only on good quality filtering or reliability evaluation. It is also of utmost importance to have thorough knowledge of what might fail and the consequences on the observed MR-spectra. Quality requirements should be defined, scanner and sequences need to be periodically and thoroughly tested, and, as a consequence of the quality cycle, better methods should be implemented/developed. The "best" solutions for each step depend on the application and the context of use; thus, what will work best for each user, may be different from one centre to the other.

As a final thought, we would like to emphasize here that besides recent efforts in the standardization of *in vivo* MRS (e.g. [Oz 2014]), there is still a *lack of* generally accepted and *implemented* standards. For MRS-methodology developers this is not a real drawback, since they are searching for improvements. However, for clinicians that want to apply the MRS-methodology, this is a real problem. Clinical routine spectroscopy requires the implementation of these standards. With this, the clinical literature would become more consistent, contributing to increase the clinical value of MRS.

5.7 Acknowledgments

The authors would like to thank Martin Pletscher for his contributions to the section on metabolite-signal cancellation due to incorrect phasing of water-suppressed signals (5.5.7), and Samuel Stettler for his contributions to the section on selection profile correction (5.5.8), as well as the authors of the articles whose materials were reused here. The authors would also like to thank Roland Kreis for the helpful discussions on the metabolite-signal cancellation artefact(5.5.7).

This work was funded by the EU Marie Curie FP7-PEOPLE-2012- ITN project TRANSACT (PITN-GA-2012-316679) and the Swiss National Science Foundation (project number 140958).

6 | Automatic Quality Control of proton MRSI data

This chapter is based on the following publication:

N. Pedrosa de Barros, R. Mckinley, U. Knecht, R. Wiest, J. Slotboom, *Automatic quality control in clinical ^{1}H MRSI of brain cancer*, NMR Biomed. 29 (2016) 563-575. doi:10.1002/nbm.3470.

6.1 Summary

MRSI grids frequently show spectra with poor quality, mainly due to the high sensitivity of MRS to field-inhomogeneities. These poor-quality spectra are prone to quantification and/or interpretation errors that can have a significant impact on the clinical use of spectroscopic data. Therefore, quality control of the spectra should always precede their clinical use. When performed manually, quality assessment of MRSI spectra is not only a tedious and time consuming task, but is also affected by human subjectivity. Consequently, automatic, fast and reliable methods for spectral quality assessment are of utmost interest. In this article, we present a new Random-Forest based method for automatic-quality-assessment of ^{1}H-MRSI brain spectra, that uses a new set of MRS signal-features. The Random-Forest classifier was trained on spectra from 40 MRSI grids that were classified as *acceptable* or *non-acceptable* by two expert spectroscopists. To account for the effects of intra-rater reliability, each spectrum was rated for quality three times by each rater. The automatic method classified these spectra with an AUC of 0.976. Furthermore, in the subset of spectra containing only the cases that were classified every time in the same way by the spectroscopists, an AUC of 0.998 was obtained. Feature importance for the classification was also evaluated. *Frequency-domain* skewness and kurtosis, as well as *time-domain* SNR in the ranges from 50 to 75 ms, and from 75 to 100 ms, were the most important features. Given that the method is able to assess a whole MRSI grid faster than a spectroscopist (∼3 seconds vs ∼3 minutes), and without loss of accuracy (agreement between classifier trained with just one session, and any of the other *labelling-session*s: 89.88%; agreement between any two *labelling-session*s: 89.03%), the authors suggest its implementation in the clinical routine. The method presented in this article was implemented in jMRUI's SpectrIm plugin.

6.2 Introduction

Localized *in vivo* Magnetic Resonance Spectroscopy (MRS) provides important biochemical information on living tissues, and has proven its added value, for example, in brain tumour diagnostics and evaluation of inborn diseases. However, *in vivo* MRS-signals are very sensitive to susceptibility differences and consequently, near bone-tissue and tissue-air interfaces, as well as when post-operative paramagnetic particles are present, artifacts are often found [Kreis 2004]. Other factors such as bone-marrow and subcutaneous lipid contamination, uncompensated

eddy currents and poor water suppression can negatively affect the quality of the spectra, and consequently reduce its clinical value. Even worse, if not aware of these phenomena, clinicians can easily draw wrong conclusions from poor quality metabolic maps, since standard quantification software does not check the quality of the spectra it quantifies.

A few metrics have been proposed to help identifying low quality spectra. Cramér-Rao Lower Bounds (CRLB)[Jiru 2006, Ratiney 2005, Yan 2002], for example, have been used for assessing the estimation error in quantification. The error is normally higher in spectra with poor quality, given that the quantification model is not able to fit well to the data. Moreover, these errors are only valid if the model applies to the measured MR-signals [Slotboom 2011], which is in general not the case for MRSI data containing artifacts. Apart from this, it was shown in a recent study [Kreis 2015] that quality filtering based on relative CRLB implies the rejection of data with low metabolite levels, and consequently, may create a bias in the estimated mean concentrations of group data, as well as prevent the clinical use of MRS in any disease leading to low metabolite levels. Other metrics, such as ER-ARSOS [Slotboom 2011, Kreis 2015, Slotboom 2009], have been proposed for estimating the error in metabolite quantification using single voxel spectroscopic (SVS) data. ER-ARSOS differs from theoretical error estimators, such as CRLB, in that it is obtained experimentally. Nevertheless, given that it requires the multiple signals acquired in the process of signal averaging, it is specific to SVS.

Manual quality checking of MRSI recordings requires much time, and can only be done by expert users. What constitutes a good quality spectrum varies between MR users, mainly depending on their level of experience and the aim for which the recording is made. In order to make MRSI more *clinically* applicable, MRSI-evaluation programs should be able to automatically recognize bad quality spectra and mask them out. This would allow a clinical user to base clinical decisions only on the reliable, high quality parts of the MRSI-maps. In this paper we introduce a random forest (RF) based method for automatic quality assessment (auto-QA) of *clinical* MRSI-data whose primary aim is to improve the clinical diagnostic process by: 1) reducing the time needed for a clinical evaluation, 2) eliminating the dependency on an experienced spectroscopist to perform the assessment, and 3) eliminating the inherently subjective aspect of manual spectral quality assessment.

Various methods for auto-QA have been reported in the past. In the work of Wright *et al.* [Wright 2008] an automatic method for assessing quality in brain tumour SVS spectra was presented, based on SVS data from the eTumour (`http://solaria.uab.es/eTumour/`) and the INTERPRET (`http://gabrmn.uab.es/interpret/`) databases. Data pre-processing was done as defined in [Tate 2006]. Manual quality control was performed by multiple spectroscopists, classifying the spectra in either *acceptable* or *non-acceptable* according to the criteria defined in [van der Graaf 2008]. The classification was done using a least-squares support vector machine classifier trained on features derived from Independent Component Analysis (ICA). The classifier had an average agreement of 87% with the expert raters. Given that the average agreement across all pairs of experts was 85%, the authors conclude that the method is at least as accurate as an expert spectroscopist.

Later on, the same classification strategy was applied to MRSI data of prostate cancer [Wright 2013]. In that work, the manual quality assessment was done by 4 expert spectroscopists, and only those spectra for which there was assessment consensus were included in the final dataset. The discarded no-consensus spectra corresponded to 17.3% of the spectra in the initial dataset. From the remaining spectra, a subset was separated and used to optimize the parameters of the classifier. Again, the obtained results allowed the conclusion that the performance of the automatic method was comparable to the expert spectroscopists: the

average agreement between pairs of experts was 88.7%, while the classifier agreed with each spectroscopist on average 87.2% of the time.

Another method for auto-QA was presented by Menze *et al.* [Menze 2008]. In this work, the freely available "R" implementation of the RF-algorithm [Liaw 2002] was used in the classification of long-TE multi-slice 2D PRESS sequence water suppressed spectra of brain tumour patients. The input features used for classification were the magnitude-spectra intensities between 0.5 ppm and 3.6 ppm. The authors compared the performance of the automatic classifier with decision rules based on SNR and Cramér-Rao-bound, showing that the first was consistently superior. The classification results obtained for the spectra *within* the PRESS box had an area under the curve (AUC) between 0.950 and 0.968, depending on whether borderline-quality spectra were considered *acceptable* or not.

The work on auto-QA presented here is also based on RF. However, a completely different strategy for selecting the input features was used. Instead of feeding the classifier with a certain range of points from the pre-processed spectra, 47 features were extracted from both time-domain (TD) and frequency-domain (FD). The relative importance of these features was then evaluated using intrinsically computed measures to the RF algorithm. Such analysis allows comparing several characteristics of the spectra in terms of their relation to the perception of spectral quality.

Another significant difference to the work of Menze *et al.* [Menze 2008] is the way in which the methods were evaluated: in this paper a cross-validation scheme was used; whereas the results of Menze *et al.*, relied in a single split of the data between train and test datasets. The advantage of using a cross-validation strategy is that, as it relies on several different divisions of the data into training and testing datasets, the obtained results should generalize better to the real application of the method in the clinical environment. As stated in [Kuhn 2013], "a test set is a single evaluation of the model and has limited ability to characterize the uncertainty in the results."

The work presented in this paper was performed in the framework of the EU-funded TRANSACT-ITN (TRANSforming Spectroscopy into a Clinical Tool: `http://www.transact-itn.eu/`), and the algorithms were implemented into a novel MRSI-analysis-GUI (SpectrIm Plug-In [Pedrosa de Barros 2014]) of the jMRUI-software (`http://www.jmrui.eu/`).

6.3 Methods

6.3.1 Ethics

All MRSI data used here was recorded from patients assigned to our institute with suspected malignant glioma, which was subsequently confirmed by histology. The measurements were performed conforming to local and national ethical regulations. All patients gave their informed consent to use the data for scientific purposes.

6.3.2 Data

All MRSI data was acquired with two different 1.5T Siemens Scanners (Aera, Avanto), using the 2D PRESS Fast MRSI product sequence with a TR of 1500 ms and water suppression (CHESS). Only long TE (135 ms) was used in this study. The 2D MRSI grid had an interpolated resolution of 32x32, with a voxel size of 5x5 mm and slice thickness of 15 mm. The excited PRESS-volume was optimized for each individual patient. In order to obtain a representative test set of spectra,

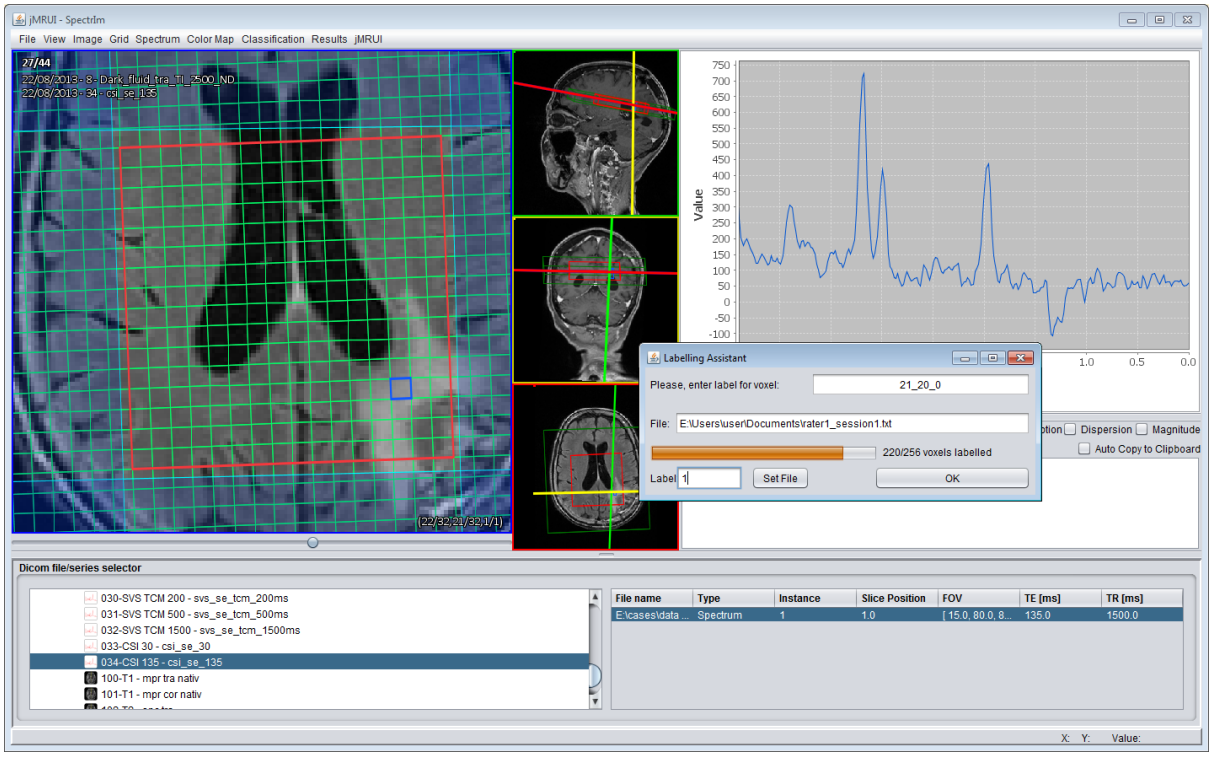

Figure 6.1: Labelling Assistant tool for jMRUI SpectrIm plugin. The assistant helps the user to label a group of spectra, one by one, into either *acceptable* or *non-acceptable*. The tool automatically iterates over all voxels displaying for each the spectrum and the localization. The figure shows the labelling-assistant requesting the user to label the selected spectrum in the tumour affected area. The tool provides information regarding the number of voxels that were already labelled by the user during the *labelling-session*. After labelling every spectrum, all features and labels are saved into a user-defined labelling file. This file can then be analysed using the Labelling Session Analyser (Figure 6.2).

a total of 12 brain tumour patients, with 2 to 4 follow-ups acquired pre- and post-operatively, were included in the study. Overall, 40 MRSI recordings with a total of 9756 spectra, all originating from *within* the PRESS box only, were used in this study. All spectra outside the excitation PRESS-box were discarded. Two different datasets were created: one where spectral features were extracted directly from "raw" non-pre-processed data –Water-Not-Removed (WNR) dataset; and a second dataset where residual-water-peak removal using jMRUI's HLSVD method was applied– Water-Removed (WR) dataset. The peak-removal was performed in a broad range, from 4.1ppm to 9.0ppm, given that poor water-suppression artifacts were often found.

6.3.3 Labelling Tools

Two new tools were developed for jMRUI's SpectrIm plugin[Pedrosa de Barros 2014]: the Labelling Assistant (Figure 6.1) and the Labelling Session Analyser (Figure 6.2). The first tool allows the user to quickly label a pre-selection of voxels from a MRSI grid, one by one, and to save these labels together with the automatically computed numerical MRSI-signal features in one file. For each voxel, the tool displays not only the spectrum but also its location, which allows incorporation of both MR-spectra and -image related information in the labelling-process. Additionally, any image series of the same MR-examination can potentially be used as background image.

Labelling a user-defined set of voxels in a MRSI grid is called a "*labelling-session*". Given

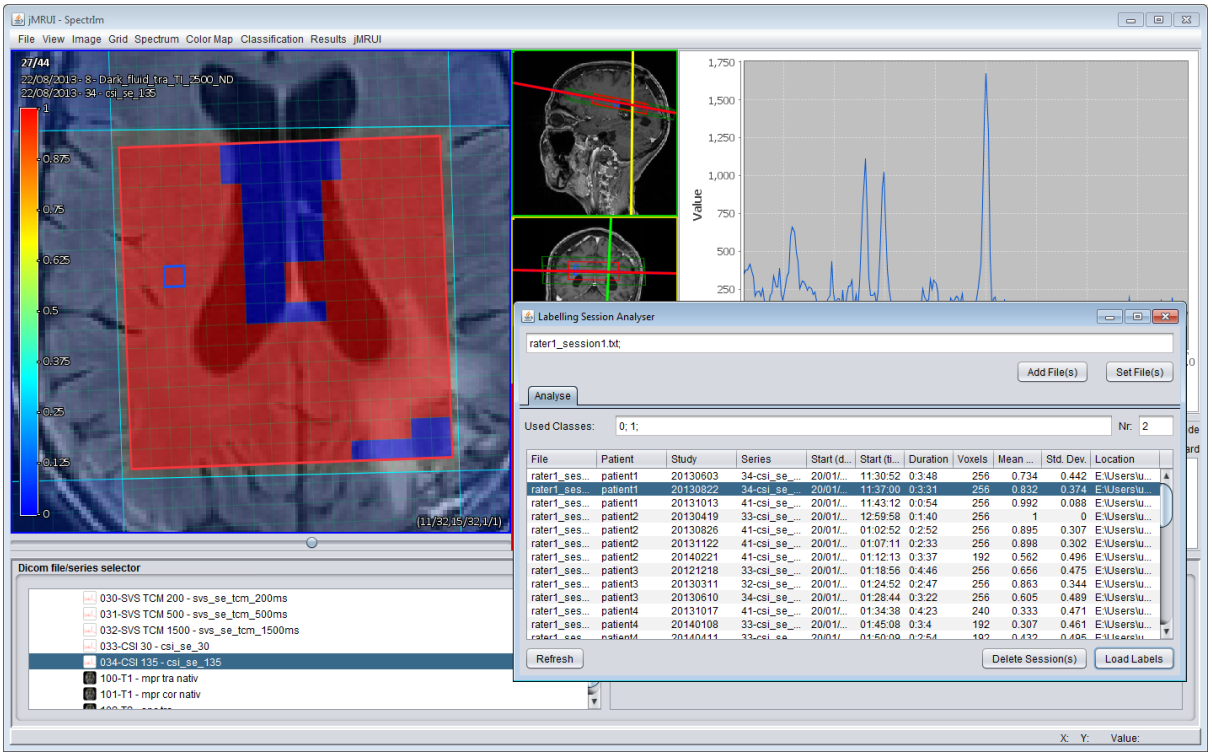

Figure 6.2: Labelling Session Analyser tool for jMRUI SpectrIm plugin that allows the user to manage *labelling-sessions*. In the screenshot, the table of this tool shows a list of several *labelling-sessions* from a certain file. In the presented case, the labels from a previous session were loaded and are being shown as a coloured map over the MRSI grid. In the map, red and blue correspond respectively to spectra which quality is *acceptable* and *non-acceptable*.

the large amount of data obtained in this way, a tool to support the management and analysis of these *labelling-sessions* was created (called Labelling Session Analyser). Using this second tool, the user is able to access the data of these sessions and view the following information: used labelling classes, patient name, study date, series number and description, labelling start date and time, labelling duration, number of classified voxels, and mean value and standard-deviation of the (numeric) labels. The user is also able to load a *labelling-session* as a label map and display it on top of the MRSI grid, as it can be seen in Figure 6.2, as well as delete a specific session from a file.

6.3.4 Manual Quality Assessment

Using the previously described tools, 2 expert spectroscopists (having 29 and 2 years of experience) labelled all the spectra 3 times each, which allowed the analysis not only of variations between spectroscopists, but also of variations between different sessions of the same spectroscopist. In this way each individual spectrum was assigned to a score between 0 (meaning 6 times classified as *non-acceptable*) and 6 (meaning 6 times labelled as *acceptable* quality).

Before labelling each grid, all residual-water resonances between 4.1 ppm and 9 ppm were removed using jMRUI's HLSVD method. After water-peak removal, automatic-phasing was performed in order to prevent bad-phasing from biasing the spectroscopists' classification. The features extracted from the MRSI-signals are such that this auto-phasing has no impact on their numerical value. No other pre-processing step was performed.

The criteria used by the spectroscopists for classifying spectra as having *non-acceptable* quality were: **a**) the presence of "ghosting" artifacts [Kreis 2004], **b**) bad shimming (frequently

occurring on localizations close to bone-tissue and tissue-air interfaces), **c**) low SNR, **d**) bone-marrow and subcutaneous lipid contaminations, **e**) signals with strongly deviating phase (even after auto-phasing), and **f**) presence of artifacts caused by post-operative paramagnetic particles (sometimes even having consequences on anatomical images, e.g. EPI).

After the manual assessment, the ground-truth was given by the *winning* labels over all sessions, following a conservative approach: a spectrum was only considered *acceptable* when classified as so more than 3 times in the 6 sessions. The authors considered as *non-acceptable* those spectra that were classified 3 times as *acceptable* and 3 times as *non-acceptable*. The reason for this is that, in the clinical environment it might be worse to have a bad-quality spectrum that was recognized as *acceptable* by the auto-QA, than a discarded good-quality spectrum. In the first situation, the bad-quality spectrum might be unnoticed and its quantification results might impact the analysis; in the later one, if the clinician realizes that relevant spectra for the analysis with *acceptable* quality are being discarded, he might still include them in the analysis but he should do it with more caution.

In some specific situations not all of the *labelling-session*s were used to determine the ground-truth. In those cases, the strategy used will be specified.

6.3.5 Spectral Features

For each spectrum, a total of 47 features were extracted. In order to avoid phase dependency, all features were extracted from the magnitude TD- and FD-signals. The choice of the features followed the following rationale: the classifier should not learn the fine details of each TD and FD signal, but perceive spectral quality from a more global perspective of these signals. This approach prevents the classifier from overfitting to specific irrelevant details of the training spectra, and is robust against spectral misalignment caused by field inhomogeneities and temperature or pH alterations, *etc*. Compared to previous approaches (e.g. [Menze 2008]), where all FD-points within a given ppm-range were used as input features, the features proposed here enable representing relevant information of a broader range of the spectra with a smaller number of features. The features used in this article can be grouped in 5 different groups:

1. *Maximum* peak SNR in given range (FD)
2. *Mean* SNR in given range (FD and TD)
3. Relative change in given range (TD)
4. Global TD features (maximum signal strength, time point of maximum signal strength, mean, standard-deviation, skewness, kurtosis)
5. Global FD features (maximum signal strength, ppm value of maximum signal strength, mean, standard-deviation, skewness, kurtosis)

The *maximum*-peak-SNR features were calculated as the SNR of the maximum value detected in a given range and will further be denoted simply by "SNR" followed by the range in ppm, e.g. "SNR [1.6; 2.2]". The *mean*-SNR features are identical to the previous ones but instead of the maximum value detected in the range, the mean value of the range is used instead. In FD, these features will be denoted by "mean SNR" followed by the range in ppm (e.g. "mean SNR[1.6; 2.2]"); and in TD by "TD mean SNR" followed by the range in *milliseconds* (e.g. "TD

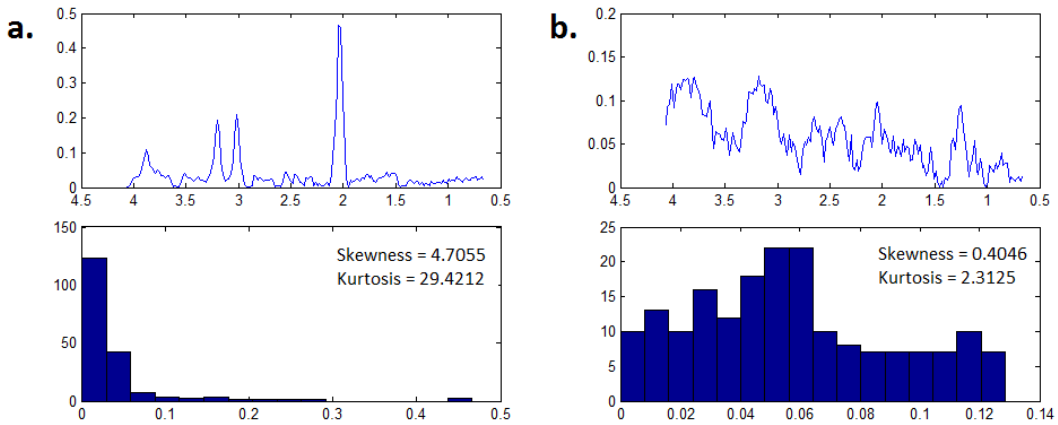

Figure 6.3: *Acceptable* (a.) and *non-acceptable* (b.) spectrum examples where one can see how spectral quality affects the FD histogram and its parameters, namely skewness and kurtosis.

mean SNR[50; 75]"). The noise-level used to calculate the SNR was computed by the standard-deviation of the signal in the 7 to 9 ppm range, for FD-features, and in the 600 to 800 ms range, for TD-features.

In FD, the *maximum*-peak-SNR and *mean*-SNR features were calculated for the following ppm ranges (and corresponding metabolites): **a.**) 0.5-1.1 (Lipids (-CH3)), **b.**) 1.1-1.6 (Lipids (-CH2)+Lactate), **c.**) 1.6-2.2 (NAA), **d.**) 2.2-2.6 (Glutamate+Glutamine), **e.**) 2.6-3.1 (Creatine), **f.**) 3.1-3.5 (Choline), **g.**) 3.5-4.1 (Creatine), and **h.**) 4.1-5.5 (residual-water/lipid). On the other hand, in TD the *mean* SNR was calculated for each of the eight ranges of 25 ms between 0 and 200 ms.

The *global mean* SNR_{0-9ppm} of the FD signals in the 0 to 9 ppm range, and the *global mean* SNR_{6-9ppm} of the 6 to 9ppm range, in which only very low-concentration brain metabolites have resonances, were also calculated. The ratio of SNR_{0-9ppm} and SNR_{6-9ppm} was computed and taken as additional feature.

The relative change in a certain range of the FID was calculated using the following formula:

$$relative\ change(k,l) = \Delta_{k,l} = \frac{s_k + s_{k+1} + s_{k+2} - s_l - s_{l-1} - s_{l-2}}{3\Delta max(s)}, l > k \tag{6.1}$$

where k and l are two points of the discrete FID, s_i is the value of the discrete FID in a point i and *max(s)* is the maximum value of the FID. The *relative change* $\Delta_{k,l}$ was calculated for the same intervals as the TD *mean*-SNR and will further be denoted by "TD Relative Change" followed by the interval (e.g. "TD Relative Change [50; 75]").

Finally, the global signal features will simply be denoted by the domain in which they were measured followed by the property being measured, e.g. "FD skewness" or "TD time max". Regarding skewness and kurtosis, given that these measures are indicators of, respectively, the histogram's symmetry and "peakedness", these features change significantly depending on whether high SNR peaks are present in the signal or not. In Figure 6.3, two examples of *acceptable* and *non-acceptable* FD spectra are shown with the corresponding histograms and values of FD skewness and kurtosis.

6.3.6 Cross-Validation and Parameter Optimization

In order to decrease the generalization error and to obtain performance measures that are independent of a particular split of the data into training and testing datasets, the classifiers were evaluated following a cross-validation scheme. Given that spectra from the same patient and MRSI grid can be highly correlated, these were not mixed between training and testing datasets. This is done to avoid artificial performance improvement. Consequently, the strategy can be described as a Leave-Patient-Out-Cross-Validation (LPOCV) scheme, where at each time the *complete* dataset of one patient was excluded from the training dataset and used as the testing dataset. Given this, for each patient p, a vector $\widehat{Y_p}$ containing all the predictions for the spectroscopic voxels of that single patient, is generated. The global performance of the method was measured using the global prediction vector $\widehat{Y_{Global}}$, defined as the concatenation of all the patient prediction vectors $\widehat{Y_p}$:

$$\widehat{Y_{Global}} = \left\{\widehat{Y_1}, \widehat{Y_2}, \ldots, \widehat{Y_N}\right\}$$

The optimization of the parameters of the classifier was performed individually for each patient, using the data of all the other patients. This step was nested in the model validation.

A more detailed description of the cross-validation scheme used can be found in Appendix A (section 6.8.1), where the strategy used to incorporate the parameter optimization into the model validation is better explained. In the same appendix, the results of the parameter optimization are also presented.

6.3.7 Random Forest

The classification algorithm used was the "R" implementation of the RF classifier [Liaw 2002, Breiman 2001, Criminisi 2013, Hastie 2001]. As the name suggests, this classifier corresponds to an *ensemble* of decision trees that are generated in a random manner. The randomness is introduced via *bagging* – i.e. each decision tree is trained with a randomly selected subset of the data – and random selection of the input features – i.e. each node of a tree uses only a predefined number of randomly selected input features (*features/node*). In this article, the Random-Forest classifiers were created with a total of 400 decision trees.

Another interesting aspect of the RF is the possibility of defining the minimum sample size for which a node may split the data into 2 child nodes (*min-node-size* [Segal 2004]). In this article, the parameters *features/node* and *min-node-size* were optimized as described in Appendix A (section 6.8.1).

The RF algorithm is able to output the probability of a given spectra belonging to a specific class (*acceptable/non-acceptable*). In this article the probability of a spectra belonging to the *acceptable* class will be denoted as *probability of acceptance*.

6.3.8 Automatic vs Manual Assessment

In order to analyse how the classifier's output compares with the manual assessment, the situation of substituting the expert's manual assessment with the predictions of the classifier was simulated. To do so, several RF-classifiers were trained with each *labelling-session* (only one session was used for each classifier). Then, the agreement between the corresponding prediction vectors $\widehat{Y_{Global}}$ and the other *labelling-sessions* that were not used in the training was measured. These results were compared with the agreement between pairs of manual *labelling-sessions*.

Table 6.1: Performance results of each individual *labelling-session*. For every session, the ground-truth was determined by the *winning* label over the other 5 sessions. The mean and standard-deviation of all session are also presented.

	spectroscopist 1			**spectroscopist 2**			**Global values**	
Session	**1**	**2**	**3**	**4**	**5**	**6**	**Mean**	**SD**
Error rate	0.081	0.081	0.082	0.107	0.096	0.098	**0.091**	0.01
Sensitivity	0.953	0.951	0.926	0.903	0.911	0.928	**0.929**	0.018
Specificity	0.852	0.853	0.901	0.873	0.889	0.848	**0.869**	0.02
PPV	0.929	0.93	0.952	0.938	0.946	0.927	**0.937**	0.009
NPV	0.898	0.893	0.853	0.81	0.825	0.85	**0.855**	0.032

6.3.9 Feature-Importance

To evaluate the relative-importance of the input features in the classification task, the *mean decrease in accuracy* after feature permutation was measured. This metric consists of the classification-error increase caused by randomly permuting the values of the feature in focus between the several data points: the bigger the error increase, the more important that feature is for the correct classification of the data.

In this article, the impact of 2 different factors on feature-importance was evaluated. Firstly, the impact of removing the residual-water resonances was analysed by comparing side-by-side the feature-importance plots of two classifiers: one trained with the WR dataset and another with the WNR dataset. Secondly, the influence of the expert that labels the spectra was evaluated. The motivation for this second comparison was the hypothesis that, besides the effort in evaluating the spectra under the same rules, each spectroscopist might be sensitive to different spectral features, which may lead to systematic differences in specific features' relative importance. To test this, the feature-importance of a classifier trained with the labels from one expert was compared with a classifier trained with the labels from the other expert. In this case the ground-truth for one spectroscopist was the *winning* label given in the 3 sessions performed by that spectroscopist (*acceptable* when classified 2 or 3 times as so and *non-acceptable* otherwise). This second test was done using the WR dataset only.

6.4 Results

6.4.1 Manual Assessment

After analysing the labels given by the two spectroscopists, we obtained the following results:

1. The overall agreement between any two *labelling-sessions* was 0.890;
2. The inter- and intra-rater rater reliability were, respectively, 0.879 and 0.908;
3. In 75.57% of the MRSI-spectra the spectroscopists gave the same label in all 6 sessions;
4. On average it took the spectroscopists 3 minutes and 8 seconds to label all spectra of *one* MRSI grid, which translates into an average required analysis-time to label one individual spectrum of 771ms;
5. The ground-truth based in all *labelling-sessions* contained 65.4% of *acceptable* spectra.

Table 6.2: Performance results obtained for each dataset. The following values are shown: area-under-the-curve (AUC), error-rate (Error), sensitivity (Sens), specificity (Spec), positive-predictive-value (PPV) and negative-predictive-value (NPV). (The detailed results obtained for each classifier of the model validation level are presented in Appendix B (section 6.8.2))

	AUC	Error	Sens.	Spec.	PPV	NPV
WR	0.976	0.075	0.943	0.891	0.942	0.892
WNR	0.948	0.113	0.912	0.838	0.914	0.834

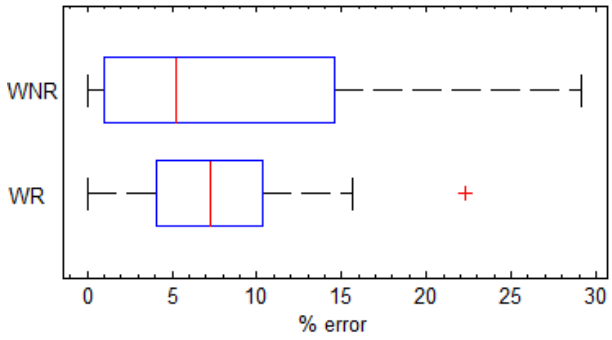

Figure 6.4: Tukey boxplots of the error-rate obtained for the several MRSI-grids. The error-rate is shown for both datasets: WR and WNR.

In Table 6.1 the performance results of each individual session are presented. For each session, the results shown consider the ground-truth as the *winning* labels over the other 5 sessions (each spectrum was considered *acceptable* when classified more than 2 times as so and *non-acceptable* otherwise). In this table we can see that – on average – each spectroscopist classified the spectra with an error-rate of 9.1%.

6.4.2 Automatic Classification

In Table 6.2, the performance results obtained with each dataset are presented. Figure 6.4 presents the Tukey boxplots of the error obtained for each MRSI-grid, for both datasets. A few examples of WR spectra for different values of *probability of acceptance*, as outputted by the classifier trained with WR data, are presented in Figure 6.5. In order to allow the comparison of the method here presented with previous work from Wright A. et al [Wright 2013], the performance of the classifier was evaluated in the subset of spectra containing all the voxels that were labelled in the same way, in every *labelling-session*. As in that work, this corresponds to analysing the method in only those spectra where there was consensus in the quality assessment. The corresponding subset of the prediction vector $\widehat{Y_{Global}}$ (n=7373) had an error-rate of 1.86% and AUC of 0.998. The results of the "automatic vs manual assessment" analysis are shown in Figure 6.6. These results show the high level of agreement between the results of the classifier and the assessment made by the experts. Noticeable is also the higher degree agreement with the experts' assessment of the classifiers trained with WR data when compared to the ones trained using the WNR dataset.

6.4.3 Spectral Features

In Figure 6.7 two feature-importance plots are shown. These plots show the effect of residual-water-peak removal on the features' relative importance (left), and the dependence on the individual spectroscopist that performed the labelling (right). In these plots are highlighted the 7 most important features for the WR dataset.

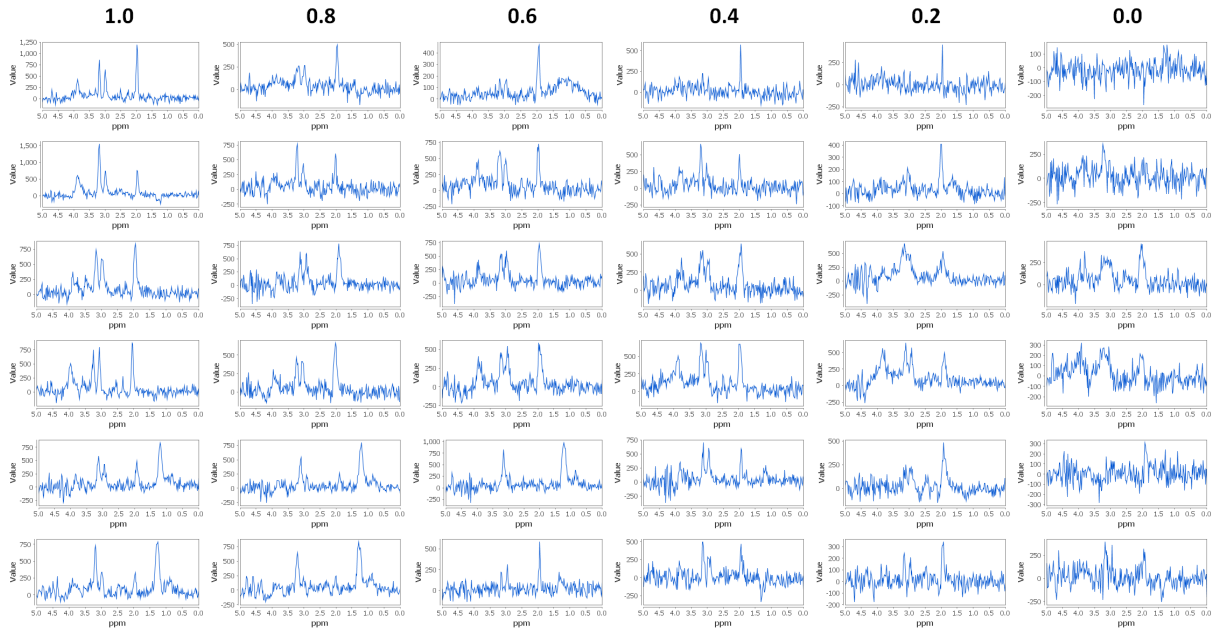

Figure 6.5: Examples of spectra from the WR dataset organized by different values of *probability of acceptance* (indicated above each column). Each column contains spectra whose quality is considered similar by the automatic-classifier.

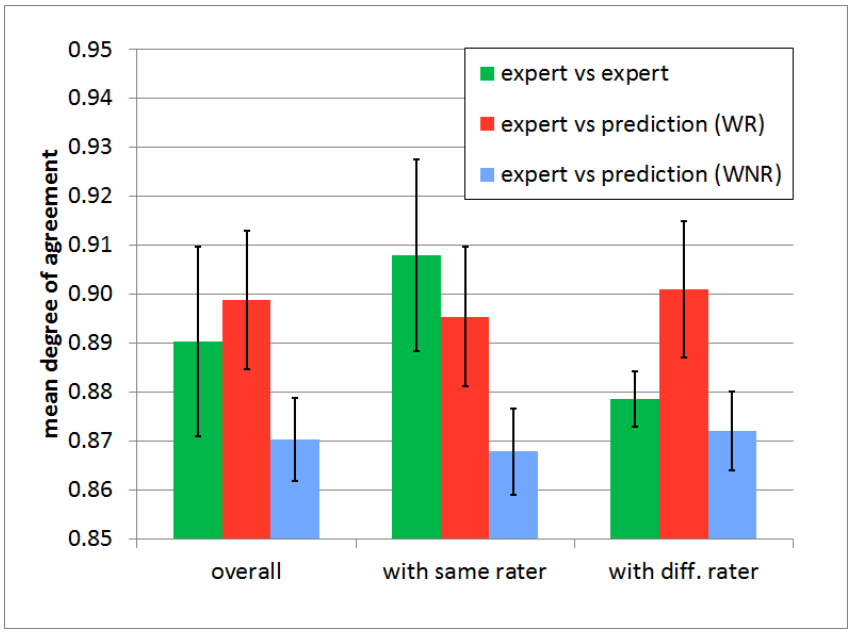

Figure 6.6: Mean degree of agreement between two *labelling-sessions* ("expert vs expert") and between the predictions from a RF classifier that was trained with the labels from one single *labelling-session*, and the remaining *labelling-sessions* ("expert vs prediction"). The agreement was measured for both the predictions made using the WR and the WNR datasets. The "overall" results represent, in the case of the "expert vs expert", the overall agreement between any two *labelling-sessions* and in the case of the "expert vs prediction" the agreement between the prediction and any other *labelling-session* that was not used for training the classifier. The results "with same rater" represent in the case of the "expert vs expert" the intra-rater reliability, and for the "expert vs prediction", the agreement between the prediction and the *labelling-sessions* from the same rater as the training data. Similarly, the results "with different rater" represent in the case of the "expert vs expert" the inter-rater reliability, and in the "expert vs prediction" case, the agreement between the prediction and the *labelling-sessions* from a different rater than the one of the training data. The error bars show the mean ± 1 standard-deviation.

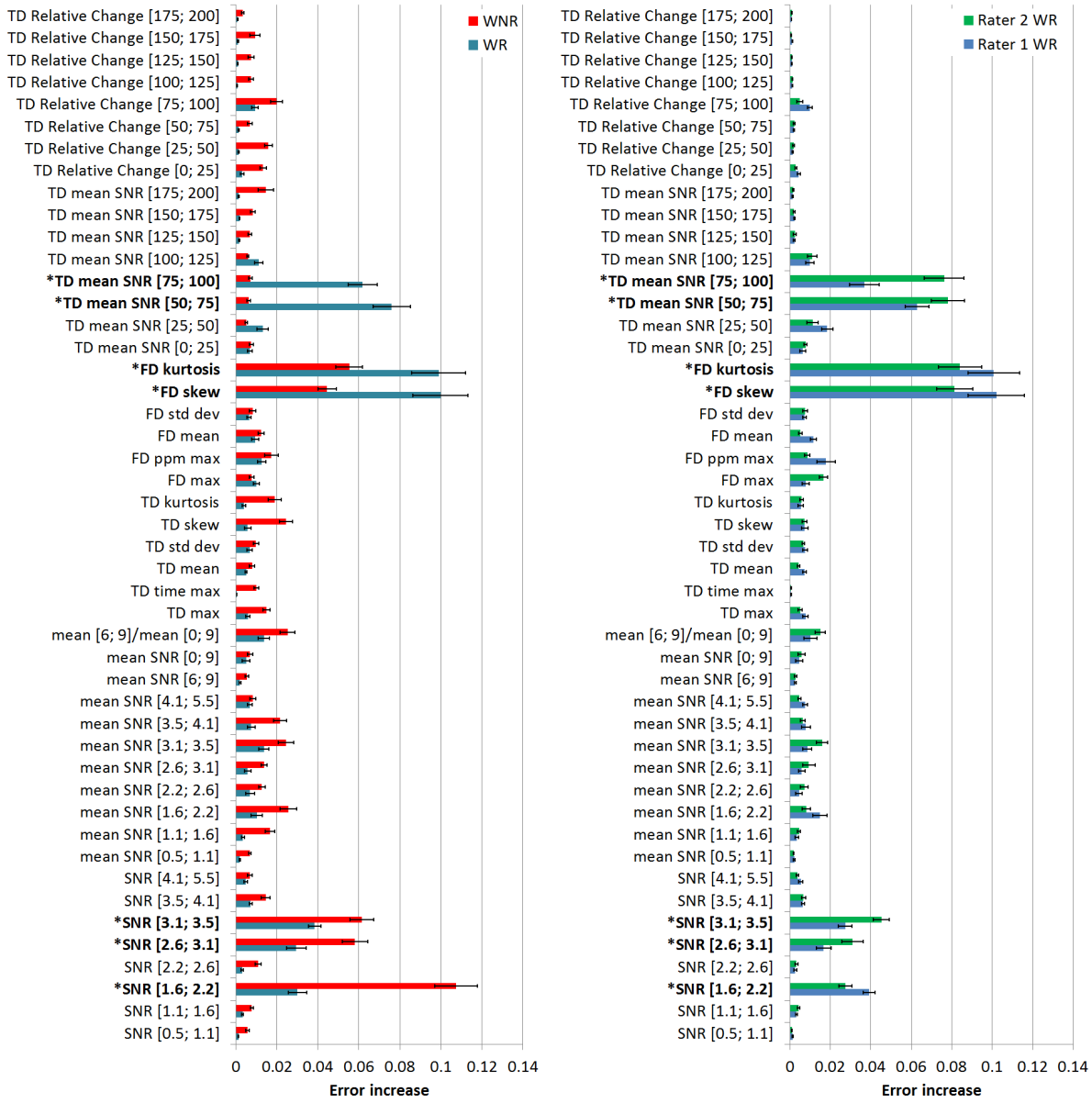

Figure 6.7: Feature importance plots. In the left plot, the feature importance of the classifiers trained using the WR dataset (blue) is compared against the classifiers trained using the WNR dataset (red). In the right plot, is shown a comparison between the feature importance of classifiers trained with class labels of two different experts. The x-axis shows the mean-error-increase observed after permuting the values of the corresponding feature in the dataset (15,20). For each one of the 4 situations presented, the mean and standard-deviation of the error increase were calculated from the 12 different random forests that are trained in the described cross-validation scheme (Classifiers from the method validation level - Appendix A (section 6.8.1)). The error bars show the mean ± 1 standard-deviation of the error increase. The 7 most important features for the WR dataset (dataset with best performance) are signalled with an *asterisk*.

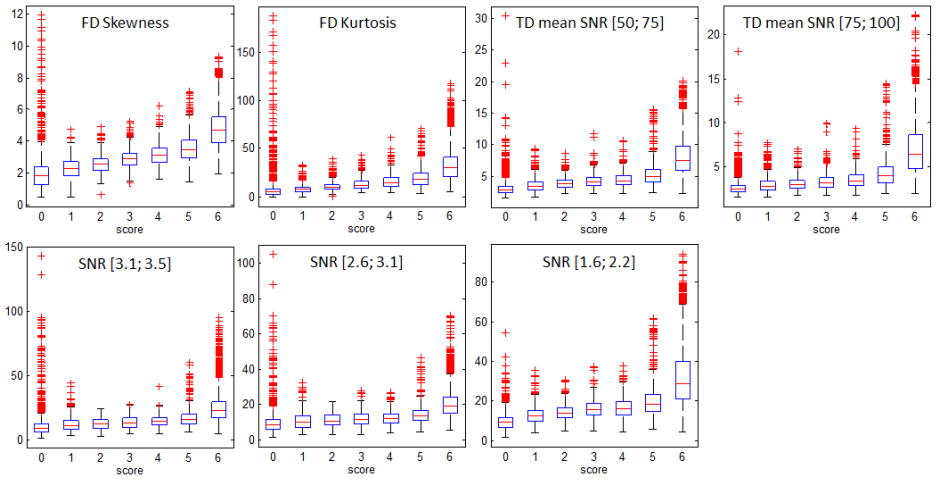

Figure 6.8: Tukey boxplots of the values of the 7 most important features of the WR dataset grouped by score (number of times classified as *acceptable* in the 6 *labelling-sessions*).

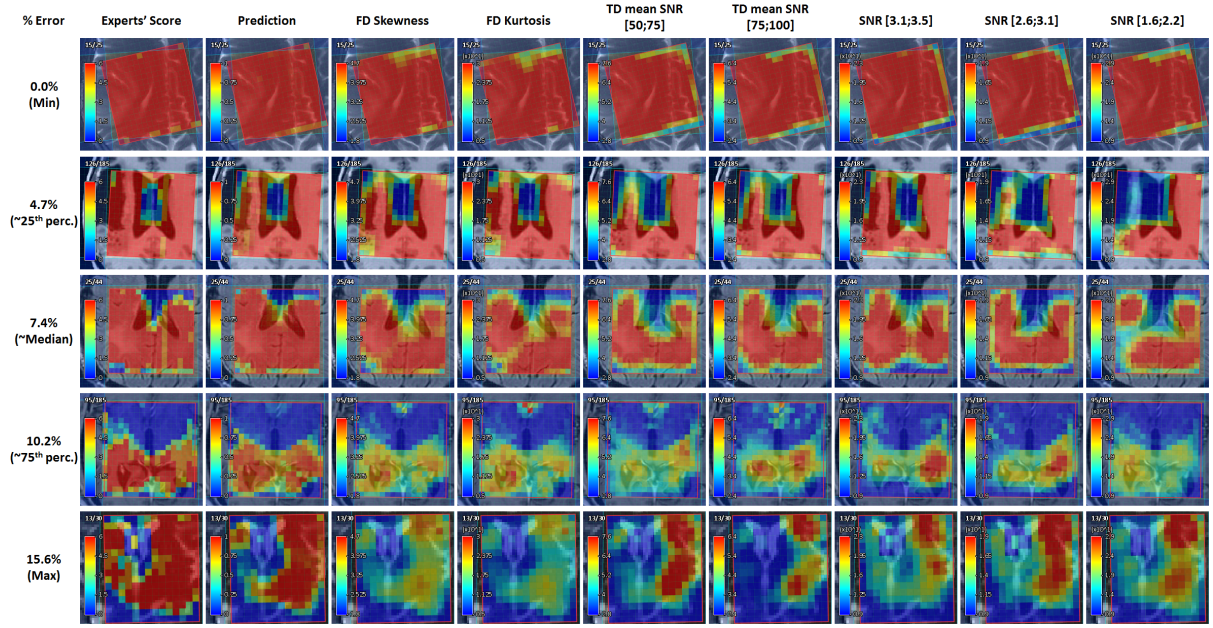

Figure 6.9: Comparison between the expert's score (score from 0 to 6, representing the number of times that each spectrum was classified as *acceptable* in the 6 *labelling-sessions*), the prediction made using the WR dataset (*probability of acceptance*), and the values for the 7 most important features (values from the WR dataset), for 5 different cases. The 5 cases were selected with error-rates close to the different marks of the corresponding boxplot (WR error-rate - Figure 6.4). Each map shows the corresponding colour scale. For each feature, the minimum of the colour scale corresponds to the median value of that feature in the spectra with score 0, and the maximum to the median value of the spectra with score 6 (see Figure 6.8). The maps were created using jMRUI's SpectrIm plugin.

The Tukey boxplots of the values of the 7 most important features (WR dataset) grouped by quality-score are shown in Figure 6.8. Then, in Figure 6.9, a comparison between the expert's score, the predictions of the classifier, and the values of the 7 most important features is shown for 5 different MRSI-grids. In order to have a representative selection of the data, the 5 cases with error-rates close to the different marks of the error-rate boxplot were selected (see Figure 6.4).

6.5 Discussion

6.5.1 Subjectivity in Manual Assessment

The results presented here show that quality-scores given by spectroscopists are subjective and not fully consistent. It was shown that the same spectra were classified in different ways by different spectroscopists. More surprising aspect is the fact that the same spectroscopist judges the same spectra in different ways. A reason for this could be that only 771ms was on average needed to classify a spectrum, and misclassifications can occur in performing the repeated task. We think that this is not the only source of variance: the subjectivity of the expert's perception of what *exactly* constitutes "spectral quality" is another factor. The auto-QA method developed here matches, to a high level of confidence, the expert knowledge and has the advantage above human-QA that it does this task in a consistent way. Because of the invariant aspect, the automated method should be recommended to prevent operator dependent bias in multi-centre clinical studies.

Another interesting aspect is the difference found in terms of the feature-importance when the classifier is trained with labels given by different spectroscopists: apparently different experts focus on different features in the spectra they are rating.

6.5.2 Automatic Classification Performance

Regarding the classifier performance, the results show that auto-QA can substitute manual-QA without loss of accuracy. For WR data, the agreement between the classifier trained with labels from just one session, and any of the other *labelling-session*s (89.88%) was equivalent to the agreement of any two *labelling-session*s (89.03%). For the classifiers trained with WNR data, the overall agreement with the expert's was slightly smaller (87.86%). It is also interesting to notice that, when WR is used, the predictions of a classifier trained with a single *labelling-session* agree more with the *labelling-sessions* of the other expert than the *labelling-session* used for training does. This might be explained by what was already mentioned, i.e. in comparison to the expert's assessment, the automatic classifier is more consistent between different assessments. In the manual case, the assessment might be biased in some situations and the variance is expected to be higher due to inconsistent assessments and possible distraction mistakes.

With an AUC of 0.976, the method here presented compares favourably to previously published methods (8,13,14). Moreover, our results were obtained performing only residual-water-peak removal and no further pre-processing. Further, in earlier published work the results were obtained using a single test-dataset, whereas here a more advanced cross-validation scheme was used (Appendix A (section 6.8.1)).

Two other referenced methods presented higher values of AUC using approaches that don't mimic the clinical use of MRS auto-QA. Firstly, in the work of Menze et al [Menze 2008], the classifier performance was computed for spectra both within and outside the PRESS

box. Since classifying noisy spectra is not very challenging, and most of the voxels are noisy voxels, this could explain the higher AUC-values which were obtained. Furthermore, since spectra outside the PRESS-box have minimal clinical interest, it is more important to compute optimized classifiers for excited MRSI voxels only. Nevertheless, a comparison with Menze et al [Menze 2008] is still possible since the results for the PRESS-box-only were also reported, namely an AUC of 0.968, which is smaller than the AUC achieved with the method presented in this paper. Secondly, in the work of Wright et al [Wright 2013] on prostate MRSI, the approach was different: *only* those spectra were included in the training of the classifier for which consensus amongst 4 spectroscopists was reached. This approach, which omits all "borderline" cases, resulted in an increased AUC of 0.983. Such strategy allows the development and testing of the method without having to deal with the critical cases, where the classification into *acceptable* or *non-acceptable* is not very clear. The classification of these "borderline" spectra is more error prone, and can vary depending on the rating spectroscopist and the exact context of the assessment. This approach is not really realistic for use in a clinical setting, since one does not find only clearly *acceptable* and clearly *non-acceptable* cases. In order to be able to compare the performance of our method with that of Wright et al, the AUC-performance in the subset of spectra that had the same label in *all* the 6 sessions was also determined. For this "consensus" subset, our method's AUC reaches a value of 0.998.

At first glance, our method seems to be better-performing than previously published methods. However, one should not forget that unambiguous comparisons can only be made if the same input data is used, and that auto-QA of spectra from different organs, regions or diseases, might pose different challenges.

Apart from the performance indicators, the continuous output of the classifier (*probability of acceptance*) shows a visible agreement with the level of spectral-quality (Figure 6.5).

6.5.3 Residual-Water-Peak Removal

In this article it was also shown that it is possible to perform auto-QA without any preprocessing even when the manual assessment is made using residual-water-removed spectra. This can be interesting in situations where the spectra need to be assessed immediately after loading the spectroscopy file. However, the gain in time-to-display the results has to be weighed against a reduction in accuracy. The difference in performance can be explained by the fact that the HLSVD-method was applied in a broad frequency offset-range and also that, in some situations, not only the residual-water-peak was removed, but also (ghosting) artifacts. These artifacts can be very prominent and, when removed, lead to a considerable increase in the quality of the corresponding spectra.

The differences in feature-importance between the classifiers trained with the WR and the WNR datasets show the impact of the residual-water-peak removal on the classification problem. For example, in the WNR case, the TD-features don't play a very important role; in contrast to the WR case in which two TD-features are amongst the most relevant ones. Another interesting aspect is the change in the relative importance of the features SNR[1.6;2.2] (NAA), SNR[2.6;3.1] (Creatine) and SNR[3.1;3.5] (Choline) between the two cases: in the WR case, these features have more or less the same importance, but in the WNR case, the feature SNR[1.6;2.2] (NAA) becomes more important than the others. This could be explained by NAA being less affected than Choline and Creatine by the variance added by the tail of the residual-water-peak. This effect reduces the spectral quality information present in the affected features. Once the residual-water-peak is removed, the effect disappears and the utility of these features is

restored.

6.5.4 Novel Features

A novel aspect of this paper is the study of the described TD- and FD-features and their relation to measured spectral quality. The RF algorithm is not only a powerful classification tool, but also very interesting for assessing the relative-importance of the features used (feature-importance). The commonly used peak-SNR features, were outclassed in feature-importance by 4 completely novel signal features that have never been used before in auto-QA, namely: FD-skewness, FD-kurtosis, TD-mean-SNR[50;75] and TD-mean- SNR[75;100]. These 4 features were the most important in the WR case, outclassing the commonly used peak-SNRs of Choline, Creatine and NAA. In Figure 6.8 the positive correlation between the value of these features and the quality-score is nicely illustrated. For the FD-skewness an almost linear relation is found. For the 7 most important features, the close relation to spectral-quality is well visible in Figure 6.9. Regarding FD skewness and kurtosis the corresponding maps are very close to the expert's score maps, confirming the close relation between their value and spectral quality. Regarding the TD-features, the TD-SNR-decrease in the range between 50 and 100ms shows a high correlation with spectral quality. The reason for this is that in this interval we have the best contrast to noise ratio: the effect of field inhomogeneity on the TD-signal strength accumulates as a function of time (both good and badly shimmed spectra start at the same amplitude, but deviate more and more until the noise level is reached).

6.5.5 Implications for Clinical Use

Finally, for the potential added-value of this QA-method in a clinical environment one can identify the following: (**1.**) In terms of its performance, it was shown that there is no significant difference between expert manual-QA and auto-QA. (**2.**) Regarding the time-needed for expert labelling (using our developed tools), a time of ∼3 minutes is needed for one MRSI grid. However, after this time investment, any new grid may be classified by the trained classifier almost instantaneously (∼3 seconds). (**3.**) If the same classifier is shared between hospitals for compatible data, this would be an important step towards standardization of QA in multi-centre studies, and could serve as basis for further classification of tissue (e.g. tumour sub-typing). (**4.**) Last but not least, this step speeds up clinical interpretation of daily routine spectroscopy in situations where an expert spectroscopist is not present. The clinician only has to analyse good quality data, which drastically simplifies the interpretation of metabolite maps: only good quality results need to be judged, and the very time intensive voxel-by-voxel visual verification step on spectral quality can be skipped, saving up to 5-10 minutes.

6.6 Conclusion

A RF based method for automatic quality control using *novel* MRS-signal features has successfully been applied to long TE MRSI-spectra. The performance of this method to distinguish good from bad quality spectra was shown to be comparable to that obtained by manual assessment. Moreover, the method massively improves the applicability by non-spectroscopic-experienced-radiologists of MRSI, since it mimics expert spectroscopist knowledge very closely. Especially for radiology centres that do not have access to expert spectroscopists, the important decision whether the MRS quality is sufficient to base a diagnosis upon, can be solved using the proposed

method. The method was implemented in jMRUI's SpectrIm plugin and, its development enabled the exploration of a novel set of MRS signal-features that show very high correlation with spectral quality. Future work should evaluate whether trained classifiers can be used to classify data acquired using different scanners and from different centres. Additionally, it will be explored whether the described spectral features can also be used for other applications, such as tumour segmentation, typing and grading.

6.7 Acknowledgements

This work was funded by the EU Marie Curie FP7-PEOPLE-2012-ITN project TRANSACT (PITN-GA-2012-316679) and the Swiss National Science Foundation (project number 140958).

6.8 Appendix

6.8.1 Appendix A: Nested cross-validation and parameter optimization

In this section the cross-validation scheme used to validate the presented method is explained in more detail. In the approach used, the parameter optimization is nested in the method validation (Figure 6.10) so that the data where the model is assessed (test sets L1) is neither included in the data used to train the classifier nor in the data used for searching the optimal classifier parameters. The cross-validation can then be seen as a 2-level Leave-one-Patient-Out-Cross-Validation (LPOCV), where the parameter optimization is a sublevel of the model validation: at each time we leave one patient out for testing the method and we use the rest of the data to find the optimal parameters (again using LPOCV) and train the classifier. The pseudo-code of the algorithm used is shown in Figure 6.11. The RF classifiers used for searching the optimal parameters contained 200 trees. The results of the parameter optimization are shown in the histograms of Figure 6.12. Based on these results, the operational versions of the classifier should be trained using the following values: 7 *features/node* and *min-node-size* equal to 4 for WR data, and 10 *features/node* and *min-node-size* equal to 1 for WNR data.

6.8.2 Appendix B: Detailed performance results

In this section we present the detailed performance results obtained for the data of each patient (Table 6.3). These results correspond to the classifiers of the method validation level (see Appendix A (section 6.8.1)).

6.8.3 Appendix C: "Probability of acceptance" threshold selection

As already mentioned, an interesting aspect of the RF classifier is its ability to provide the probability of a given spectra belonging to the acceptable class (probability of acceptance). Given the continuous nature of this output, it is possible to change the threshold that separates acceptable and non-acceptable spectra, which can be seen as defining stricter or less strict quality criteria. This has direct impact on the sensitivity, positive predictive value (PPV) and accuracy (Figure 6.13). Changing this threshold might be interesting, for example, in some applications that are more sensitive to bad-quality spectra. In such applications it could be wise to increase the threshold, sacrificing some of the sensitivity for a higher PPV. On the other hand, when the data is scarce and the methods that are being used are not so sensitive to bad-quality,

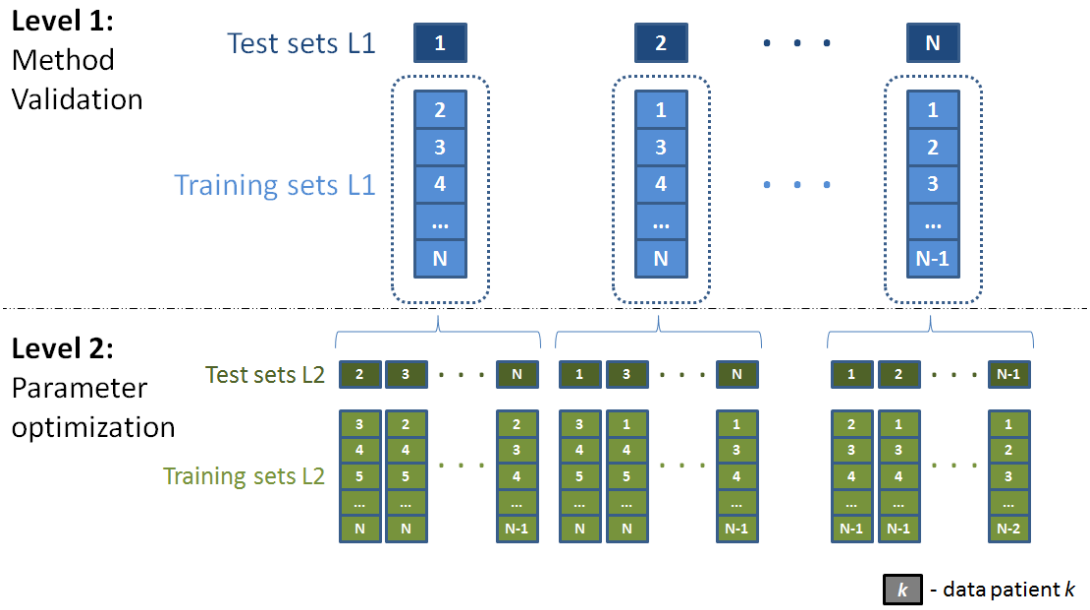

Figure 6.10: Organization of the data used in the cross-validation scheme used. Here each numbered cube represents the data of one patient. As it can be seen the parameter optimization is a sublevel of the method validation. The data is initially split at the first level into a test set L1, containing the data of one patient, and a training set L1, containing the data of all the other patients. The training set L1 is then used for the parameter optimization and sub-divided into level-2 test sets and training sets. This subdivision is again done by removing at each time the data of one patient for testing (test sets L2) and using the remaining patients for training (training sets L2).

Algorithm

for each patient i {
$testSet_{L1} = \{X(i), Y(i)\}$
$trainingSet_{L1} = \{\mathbf{X} \setminus X(i), \mathbf{Y} \setminus Y(i)\}$
for each patient $j \neq i$ {
$testSet_{L2} = \{X(j), Y(j)\}$
$trainingSet_{L2} = trainingSet_{L1} \setminus \{X(j), Y(j)\}$
for each parameter set $\theta_k \in \Theta$ {
$RF' = train(\theta_k, trainingSet_{L_2})$
$\widehat{Y'_{\theta_k}}(j) = test(RF', testSet_{L_2})$
}
}
$\theta_i = \arg\max_{\theta_k \in \Theta}(AUC(\mathbf{Y} \setminus Y(i), \bigcup_{j \neq i} \widehat{Y'_{\theta_k}}(j)))$
$RF_i = train(trainingSet_{L_1}, \theta_i)$
$\widehat{Y_{\theta_i}}(i) = test(RF_i, testSet_{L_1})$
}
$\widehat{Y_{Global}} = \bigcup \widehat{Y_{\theta_i}}(i)$

Figure 6.11: Pseudo-code of the algorithm used to perform the Nested Cross-Validation.

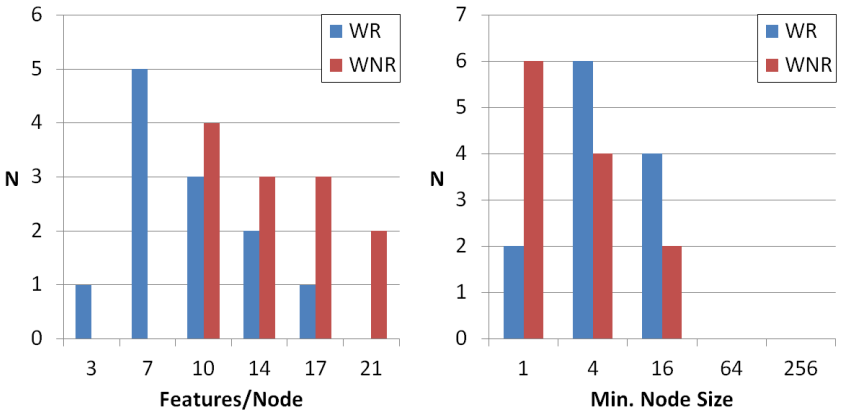

Figure 6.12: Histograms of the best parameter values obtained for the several data subsets. The results for both the WR and WNR datasets are shown.

Table 6.3: Performance results obtained for the predictions made for each patient, as well as for the global prediction vector. The results were obtained according to the Nested CV scheme described, and are shown for both the WR and the WNR datasets. The following values are shown: number of voxels, percentage of acceptable voxels (%Accept), area-under-the-curve (AUC), accuracy (Acc.), sensitivity (Sens.), specificity (Spec.), positive-predictive-value (PPV) and negative-predictive-value (NPV).

			WR						WNR					
Patient	**voxels**	**%Accept**	**AUC**	**Acc.**	**Sens.**	**Spec.**	**PPV**	**NPV**	**AUC**	**Acc.**	**Sens.**	**Spec.**	**PPV**	**NPV**
1	768	0.848	0.997	0.982	0.994	0.915	0.985	0.964	0.996	0.974	0.988	0.897	0.982	0.929
2	960	0.861	0.977	0.954	0.977	0.812	0.970	0.850	0.967	0.940	0.955	0.842	0.974	0.752
3	768	0.682	0.965	0.922	0.956	0.848	0.931	0.900	0.930	0.884	0.954	0.734	0.885	0.882
4	624	0.311	0.967	0.913	0.856	0.940	0.865	0.935	0.928	0.854	0.825	0.867	0.737	0.916
5	736	0.781	0.969	0.916	0.943	0.820	0.949	0.800	0.946	0.851	0.863	0.807	0.941	0.622
6	1108	0.663	0.975	0.919	0.944	0.869	0.934	0.888	0.964	0.898	0.898	0.898	0.946	0.817
7	716	0.789	0.974	0.933	0.954	0.854	0.961	0.832	0.942	0.897	0.935	0.755	0.935	0.755
8	668	0.352	0.979	0.940	0.928	0.947	0.905	0.960	0.929	0.894	0.762	0.965	0.923	0.882
9	964	0.804	0.986	0.949	0.956	0.921	0.980	0.837	0.973	0.920	0.925	0.899	0.974	0.746
10	540	0.422	0.948	0.878	0.882	0.875	0.838	0.910	0.957	0.904	0.829	0.958	0.936	0.885
11	1024	0.610	0.958	0.880	0.891	0.862	0.910	0.835	0.938	0.838	0.872	0.784	0.864	0.796
12	880	0.513	0.957	0.907	0.905	0.909	0.913	0.901	0.843	0.790	0.925	0.648	0.734	0.891
Global	**9756**	**0.654**	**0.976**	**0.925**	**0.943**	**0.891**	**0.942**	**0.892**	**0.948**	**0.887**	**0.912**	**0.838**	**0.914**	**0.834**

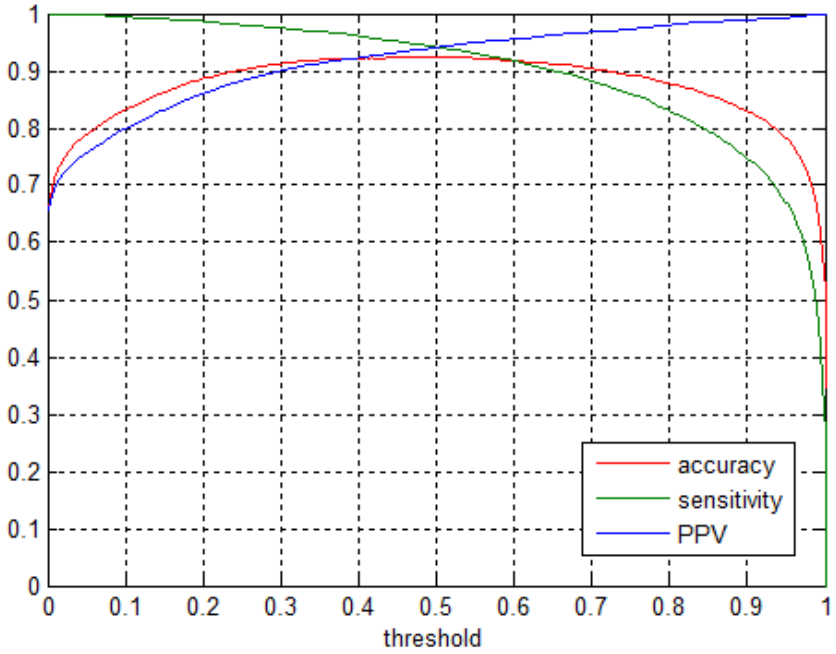

Figure 6.13: Sensitivity, PPV and accuracy of the classifier depending on the *probability of acceptance* threshold. The results are shown for the WR dataset only.

it might make sense to lower the threshold. The ability to select the spectra from an MRSI grid that have a probability of acceptance higher than a user-defined threshold was implemented in SpectrIm.

7 | Improving labeling efficiency in automatic quality control of MRSI data

This chapter is based on the following material:

N. Pedrosa de Barros, R. Mckinley, R. Wiest, J. Slotboom, *Improving labeling efficiency in automatic quality control of MRSI data*, Magn. Reson. Med. (2017). doi:10.1002/mrm.26618.

7.1 Abstract

Purpose: To improve the efficiency of the labeling task in automatic quality control of MRSI data.

Methods: 28'432 short and long-TE spectra (1.5T, PRESS, TR=1500ms), from 18 different brain-tumour patients, were labeled by two experts as either "accept" or "reject", depending on their quality. For each spectrum, 47 signal features were extracted. The data was then used to run several simulations and test an active learning approach using uncertainty sampling. The performance of the classifiers was evaluated as a function of the number of patients in the training-set, number of spectra in the training-set, and a parameter α used to control the level of classification uncertainty required for a new spectrum to be selected for labeling.

Results: The results showed that the proposed strategy allows reductions of up to 72.97% for short-TE and 62.09% for long-TE in the amount of data that needs to be labeled, without significant impact in classification accuracy. Further reductions are possible with significant but minimal impact in performance.

Conclusion: Active learning using uncertainty sampling is an effective way to increase the labeling efficiency for training automatic quality control classifiers.

7.2 Introduction

Bad quality spectra are commonly observed in in-vivo Magnetic Resonance Spectroscopy. Multiple factors can cause MRS(I) signal artifacts, a topic that is extensively covered in the work of Kreis ([Kreis 2004]). Bad quality spectra may prevent proper spectral quantification, which on its turn may lead to misinterpretation of the spectra. In order to guarantee proper quantification, quality control strategies are needed. For quality control of MR spectra it is important to identify signal artifacts that may corrupt the quantification first, and then to evaluate the quality of the quantification ([Slotboom 2009]) and the associated error ([Cavassila 2001, Jiru 2006, Kreis 2015, Slotboom 2011, Bolliger 2013, Ratiney 2005]). For the

identification of artifacts, simple visual inspection is generally performed, nevertheless this is only possible for single voxel acquisitions and when an expert is available. In the case of multi-voxel acquisitions, thousands of spectra are usually generated in just one acquisition, making the visual inspection of all the data a Sisyphean task. In order to automatize artifact and poor SNR detection in MRSI, a few machine-learning based methods have been proposed ([Menze 2008, Wright 2013, Pedrosa de Barros 2016b, Pedrosa de Barros 2016a]). Such methods allow to check thousands of spectra in a matter of seconds with a performance that is identical to manual inspection ([Pedrosa de Barros 2016b, Pedrosa de Barros 2016a]) performed by human MRS experts. Besides the obvious time-saving potential of these methods, a considerable amount of time is needed for manually labeling the data that is used for training the automatic classifiers.

The selection of the training samples is often done by identifying a "reasonable" amount of MRSI acquisitions and by labeling all the corresponding spectra. However, not every individual spectrum provides relevant information to the learning algorithm, and the dependency that the training sample size has on the classifier performance is often not analyzed.

Labeling efficiency can be increased with the use of active learning approaches ([Debarr 2009, Patra 2012, Maiora 2014, Cohn 1996, Tuia 2009, Silva 2004, Kutsuna 2012]) where the learning method is able to actively select the subset of examples that need to be labeled by the user. This is an iterative process where 1) the algorithm selects a subset of examples, 2) the user labels these selected examples and 3) the learning method is updated with the new data before selecting again more data for labeling.

Uncertainty sampling ([Lewis 1994]) is a sampling strategy used in active learning that gives priority to the selection of those examples whose class membership is more uncertain. Since these cases are usually close to the class separation boundaries, they contain most of the information needed to separate the different classes. In different applications, uncertainty sampling has shown to improve the effectiveness of the labeling procedure drastically ([Patra 2012, Maiora 2014, Lewis 1994]). In this work, we've investigated the use of active learning with uncertainty sampling for automatic quality control of MRSI data with the objective of increasing manual labeling efficiency. The proposed learning strategy may contribute to a faster and easier implementation of automatic quality control methods.

Several simulations were run to analyze the effect of the proposed strategy on the reduction of the labeled sample size and on the performance of the classifiers. Before running the simulations, 2 experts labeled all spectra from a long-TE and a short-TE datasets. To simulate the partial labeling of the data in each simulation, only the examples selected for labeling were added to the training-set. This allowed to test different sampling strategies and evaluate the dependency of the classification performance on the training-set size.

7.3 Methods

7.3.1 Data acquisition

Short (30 ms) and long (135 ms) TE data were acquired on 1.5T MR-scanners (Siemens Aera, Avanto) using the PRESS localization method, CHESS water-suppression, TR=1500 ms, and a 32x32 grid (interpolated from 12x12 k-space samples). A total of 28'432 spectra were included in the study, 14'216 for each TE. These spectra were all selected from within the PRESS-box only, and all other spectra from outside the selected volume were discarded. The spectra were retrieved from 58 different studies of 18 brain-tumour patients, performed pre- and

post-operatively, and where short and long TE data were acquired sequentially in the same localization.

The measurements were performed conforming to local and national ethical regulations, meaning that all patients gave their informed consent to use the data for scientific purposes.

7.3.2 Data processing and feature extraction

All data processing, feature extraction and labeling were performed using the SpectrIm plugin of jMRUI ([Pedrosa de Barros 2014], `www.jmrui.eu`). Residual-water-peak-removal was performed using HLSVD prior to feature extraction. A total of 47 features were extracted from time-domain (TD) and frequency-domain (FD) magnitude spectra, as in ([Pedrosa de Barros 2016b, Pedrosa de Barros 2016a]). The following groups of features were used:

- Local maximum peak SNR (FD)
- Local mean SNR (FD and TD)
- Local relative change (TD)
- Global TD features (maximum signal strength, time point of maximum signal strength, mean, standard-deviation, skewness, kurtosis)
- Global FD features (maximum signal strength, ppm value of maximum signal strength, mean, standard-deviation, skewness, kurtosis)

A detailed description of the features used can be found in ([Pedrosa de Barros 2016b, Pedrosa de Barros 2016a]).

7.3.3 Labeling

The spectra were manually labeled by two expert spectroscopists (further referred to as raters) based on their quality as either "accept" or "reject". The raters could visualize both the spectra and the voxel localization during the quality labeling process. The features for rejecting spectra included: "ghosting" artifacts, bad-shimming, low-SNR, lipid-contaminations, strongly deviating phase, and post-operative-derived artifacts. First, the labeling was performed independently by each rater. In a second phase, a consensus decision was made for all those spectra for which there was disagreement between the raters. The resulting consensus-labels constituted the ground-truth used for training and testing the automatic classifiers. For each individual session, the time spent for labeling each MRSI-grid was recorded.

7.3.4 Random-forest

A random-forest ([Liaw 2002, Breiman 2001]) classifier ("R" implementation, 200 trees, maximum depth, 7 features per node) was used for the automatic assessment. This classifier provides for each test example an approximation to the class membership probabilities, calculated as the % of trees of the ensemble that vote for each class. Once these probabilities are known, a class can be assigned to each example. Consequently, in a binary-class problem the decision threshold is often selected as 0.5. Despite this, in the problems characterized by a considerable class-imbalance (i.e. the number of members differ considerably between classes), this decision

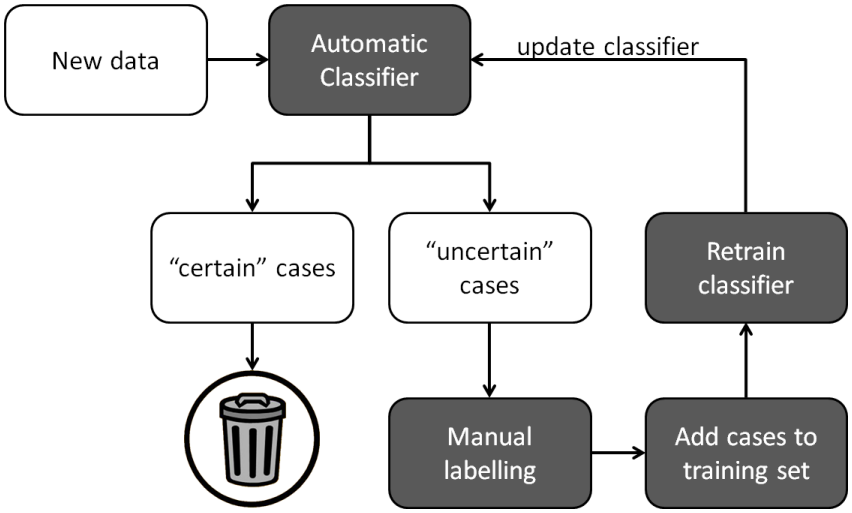

Figure 7.1: Proposed active learning strategy for training the automatic classifier where only a subsample, containing the examples with more uncertain class membership, is labeled and included in the training-set.

boundary might lead to poor sensitivity or specificity. Therefore, in those cases, a different decision threshold is often used. The examples with the highest classification uncertainty have class membership probabilities closer to the decision boundary. Thus, these can be selected by defining a range of probabilities centred on the boundary. In this study, considering that we have a binary-class problem with reasonably balanced classes([Pedrosa de Barros 2016b]), the most uncertain cases were identified as the examples that satisfy:

$$-\alpha < p(y = "accept"|x) - 0.5 \leq \alpha \tag{7.1}$$

where the parameter α is used to control the minimum level of uncertainty of the selected sample. In this study, the effect of α on the number of spectra that is selected for labeling, and on the classification performance was analyzed.

7.3.5 Active learning

The active learning strategy proposed here is depicted in Figure 7.1. As can be seen in the workflow, for each "block" of new data, the previously trained classifier is used to determine the subset of uncertain examples that will be manually labeled and included in the training-set. All the cases for which the classifier is more certain regarding their class membership are rejected. In this work, short and long TE were handled separately and one "block" corresponds to all the spectra from all MRSI acquisitions of a single patient, collected pre- and post-operatively.

7.3.6 Simulations

In order to validate the proposed active-learning method, several simulations were performed. Each simulation consisted of several iterations, where in each iteration the spectra from a single patient was analyzed and the uncertain examples were added to the training-set. The performance was measured at the end of each iteration after adding the new spectra to the training-set and retraining the classifier. In each iteration, the labeling of the selected uncertain examples was performed by drawing the required labels from the completely labeled dataset containing all 18 patients (see labeling section). Simulations using short and long TE data were performed separately.

In each simulation, the performance was evaluated on a validation-set corresponding to the data of one patient that was excluded from the iterative process and consequently never added to the training-set (Leave One Out Cross Validation - LOOCV). Given that in each iteration a previously trained classifier is required to select the spectra for labeling, the initial classifier was trained using all spectra from the first patient (see pseudo-code, lines 10 and 11). Another aspect that was considered was the order in which the patients were iterated over. It is expected that data from different patients contains different levels of information and that the evolution of the performance in each iteration might depend on the order in which the patients are included. Consequently, the patients' order was shuffled in each simulation, and for each set of parameters (TE, α and patient selected for validation-set) the simulation was repeated 50 times. Five different values of α were tested: 0.1, 0.2, 0.3, 0.4 and 0.5, representing 0.5 the case where, in each iteration, all spectra from the corresponding patient are labeled and added to the training-set.

The pseudo-code corresponding to the simulations performed is presented below in algorithm 3.

Algorithm 3 Simulations' pseudo-code

1: **for** TE $t \in \{30ms, 135ms\}$ **do**
2: **for all** $\alpha \in \{0.1, 0.2, 0.3, 0.4, 0.5\}$ **do**
3: **for** $n = 1$ **to** 50 **do**
4: **for all** patient $i \in \{1, \ldots, N\}$ **do**
5: $ValidationSet \leftarrow DataSet(patient = i, \text{TE}= t)$
6: $TrainingCandidatesSet \leftarrow DataSet(patient \neq i, \text{TE}= t)$
7: $PatientIndexes \leftarrow Shuffle(\{1, \ldots, N\} \setminus \{i\})$
8: **for** iteration $j = 1$ **to** $N - 1$ **do**
9: $index \leftarrow PatientIndexes(j)$
10: **if** $isEmpty(TrainingSet)$ **then**
11: $TrainingSet \leftarrow TrainingCandidatesSet(patient = index)$
12: **else**
13: **for all** spectrum $k \in TrainingCandidatesSet(patient = index)$ **do**
14: $p(y_k = "accept"|x_k) \leftarrow Test(RF_{j-1}, x_k)$
15: **if** $-\alpha < p(y_k = "accept"|x_k) - 0.5 \leq \alpha$ **then**
16: Add (x_k, y_k) to $TrainingSet$
17: **end if**
18: **end for**
19: **end if**
20: $RF_j \leftarrow Train(TrainingSet)$
21: $\hat{Y}_{i,j} \leftarrow Test(RF_j, X_i \in ValidationSet)$
22: $Performance_{t,\alpha,n,i,j} \leftarrow Performance(\hat{Y}_{i,j}, Y_i \in ValidationSet)$
23: **end for**
24: **end for**
25: **end for**
26: **end for**
27: **end for**

The performance measures recorded were: Accuracy, area-under-the-curve (AUC), and number of spectra and % of spectra with label "accept" (%*accept*) in the training-set (see line

22). The %*accept* metric was used to evaluate the class balance of the training-set. For each TE and α value, the mean cross-validated accuracy, area-under-the-curve (AUC), number of spectra and %*accept* were calculated in function of the number of patients added to the training-set.

7.4 Results

7.4.1 Labeling

The average agreement between the raters' individual assessments was 85.80% for short-TE and 89.05% for long-TE data. The percentage of spectra labeled, after consensus, as "accept" was 71.55% for short-TE and 74.76% for long-TE data. Regarding the time spent for labeling, each rater required an average of 0.96 seconds to label one single spectrum during the individual assessment. This represents the average time that a rater requires for attributing a label to each spectrum, and does not include the time spent in other activities, such as loading data, removing the residual water or the time which was needed to reach a consensus decision.

7.4.2 Classification Uncertainty and class-separation

Figure 7.2 shows a scatter plot of two of the most important features for the discrimination between acceptable and non-acceptable spectra: "frequency domain (FD) skewness" and, "time domain (TD) mean SNR" in the range of 50 to 75ms after pulse echo maximum. These features were selected based on the results concerning feature importance presented in ([Pedrosa de Barros 2016b, Pedrosa de Barros 2016a]). Besides only using two features, the plot shows a good separation between classes. The figure shows that spectra with more uncertain class membership (darker) are located closer to the boundary between the two groups, confirming that the information needed to determine the class separation boundaries is contained in the more uncertain spectra.

7.4.3 Performance impact

Figure 7.3 shows the mean cross-validated AUC, and Figure 7.4 the classification accuracy depending on the number of patients (left) and the number of spectra in the training-set (right). The comparison between the complete-labeling case (α=0.5) and the other cases shows the increased efficiency that uncertainty sampling can add to the labeling task and the minor impact that it has on the performance of the classifiers, namely for α between 0.2 and 0.4. The results show also the clear differences in performance between short and long-TE data classification, and that AUC changes more with α than accuracy.

The analysis of the performance differences between the α=0.5 (complete-labeling) and each other α value for the classifiers trained with data from all patients but the one used for validation, is presented in Figure 5. In these boxplots, we see that for certain values of α there are no statistical significant differences (two-tailed t-test, unequal variance) between the classifiers trained with the complete dataset and the ones trained with just the uncertain examples. For the other cases the differences are significant, but minor.

The maps showing the predictions made by classifiers trained with different α, for 4 different MRSI grids (2 examples with both short and long TE acquisitions) are presented in Figure 7.6. Besides the overall good match between the ground-truth and the predictions, we see a relation between α and the certainty of the predictions: classifiers trained using a smaller α produced generally more uncertain predictions, namely for bad-quality spectra.

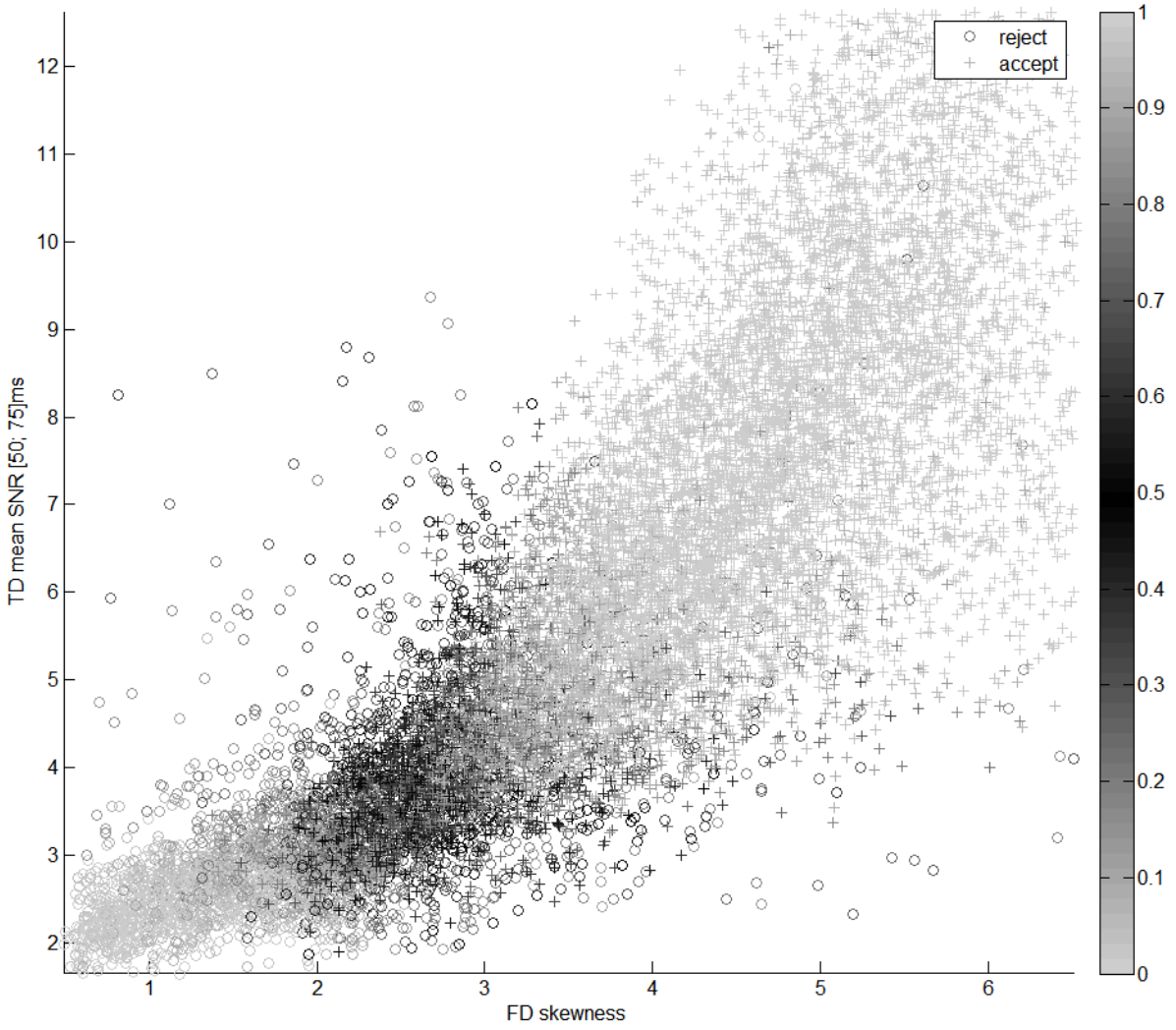

Figure 7.2: Scatter plot showing all the 14216 long TE spectra. The spectra can be distinguished based on their label (symbol used) and $p(y = "accept"|x)$ (gray scale). The more uncertain cases, with class probabilities closer to 0.5, were colored darker. The two selected features (time domain mean SNR in the range between 50 and 75 ms, and frequency domain skewness) are among the ones with the highest correlation with the quality of long-TE spectra, as shown in previous work ([Pedrosa de Barros 2016b, Pedrosa de Barros 2016a]). The plot shows that the cases with more class membership uncertainty are found near the class separation boundary. For sake of simplicity, only 2 features were chosen and only long-TE data is represented.

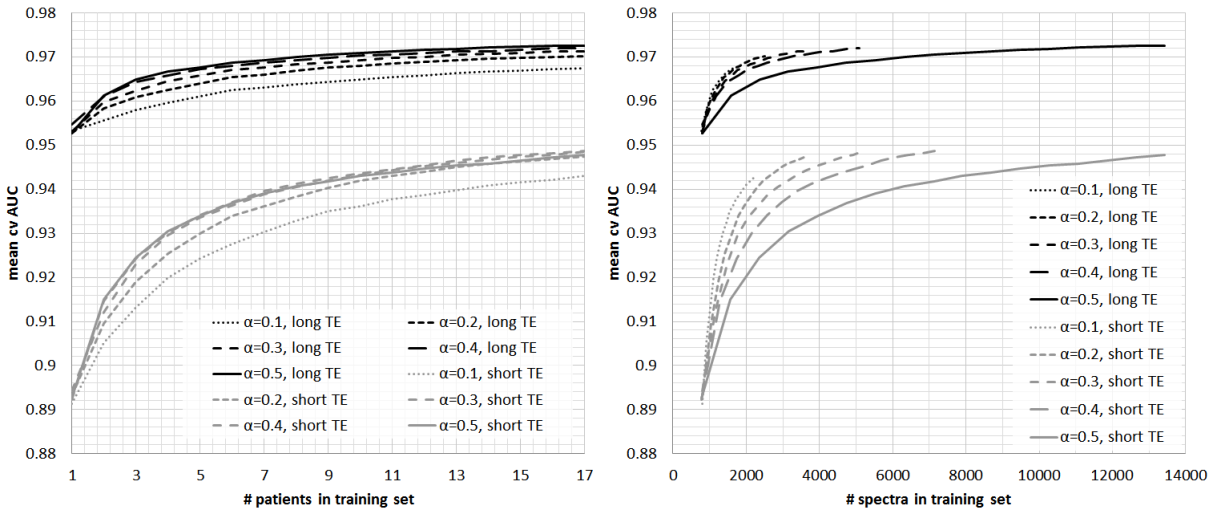

Figure 7.3: Mean cross-validated AUC as a function of the number of patients (left) and number of spectra (right) added to the training-set, for both short and long TE and for different values of α.

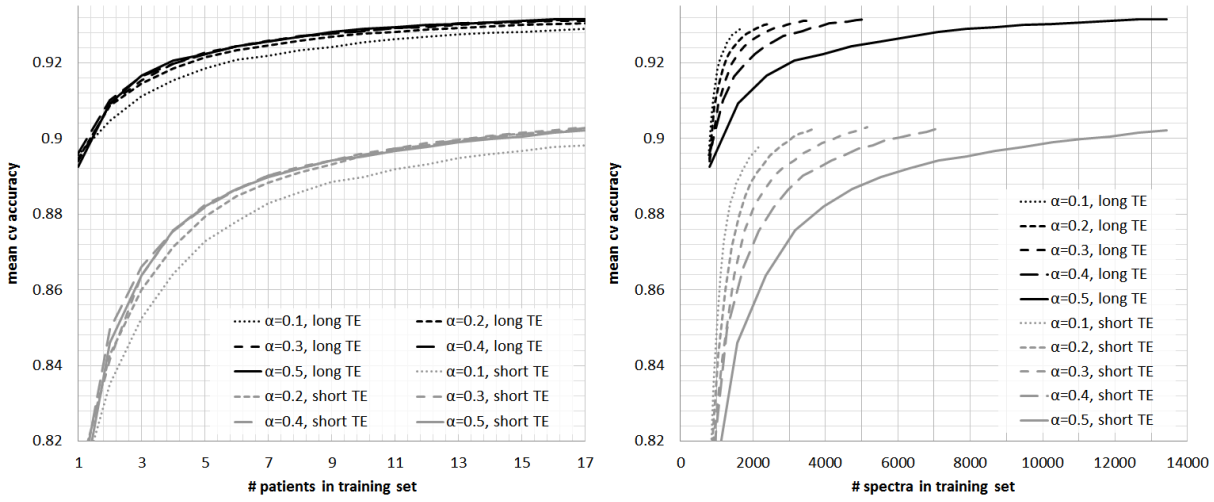

Figure 7.4: Mean cross-validated accuracy as function of the number of patients (left) and number of spectra (right) added to the training-set, for both short and long TE and for different values of α.

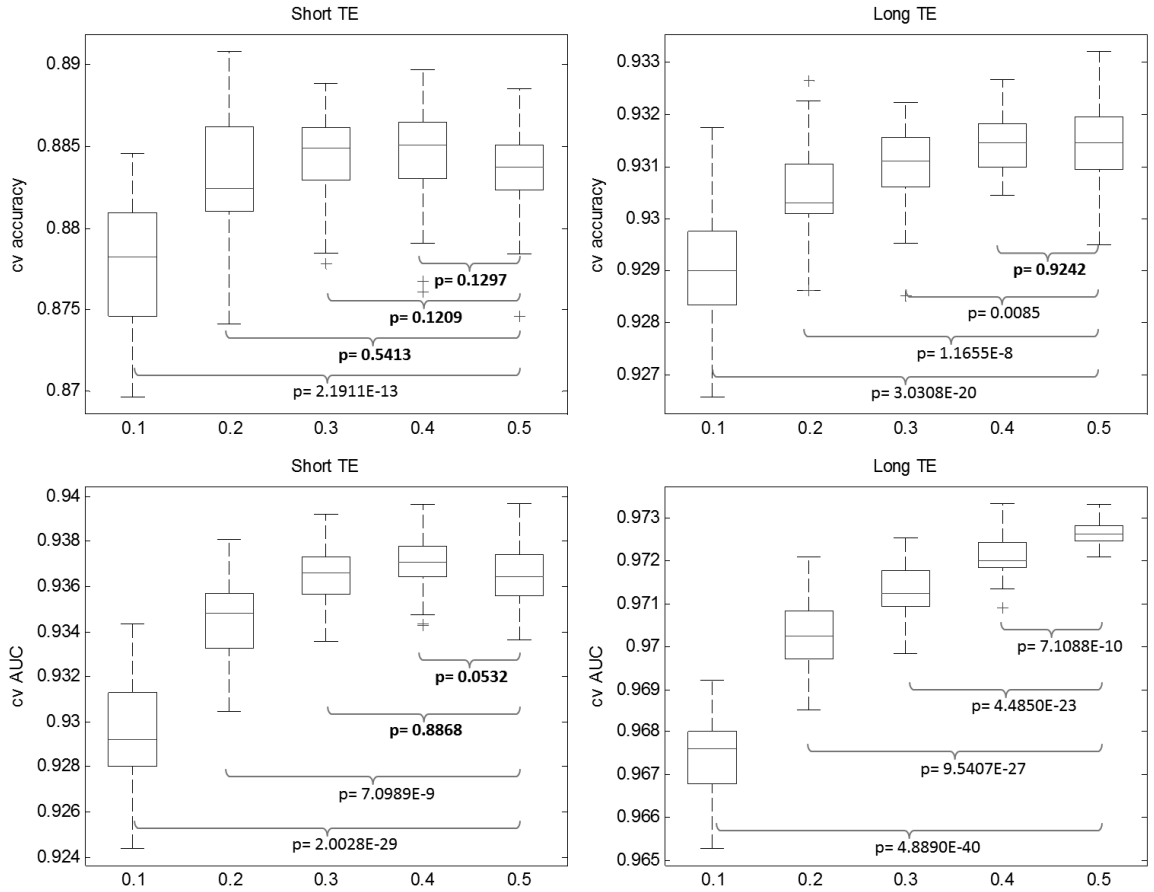

Figure 7.5: Boxplots of the cross-validated accuracy and AUC for short and long TE data and for different values of α. The performance measures were calculated using the classifiers trained with data from all patients (but the one excluded for testing). The boxplots show how the performance changed between different repetitions of LOOCV. The differences in performance between different repetitions are mainly caused by the differences in the order in which the patients are iterated over, but also by the random aspect of the classifier used (random forest). The statistical significance of the differences, between the accuracy and AUC of $\alpha = 0.5$ and the other α values, was computed using a two-tailed t-test considering 2 distinct populations with unequal variance.

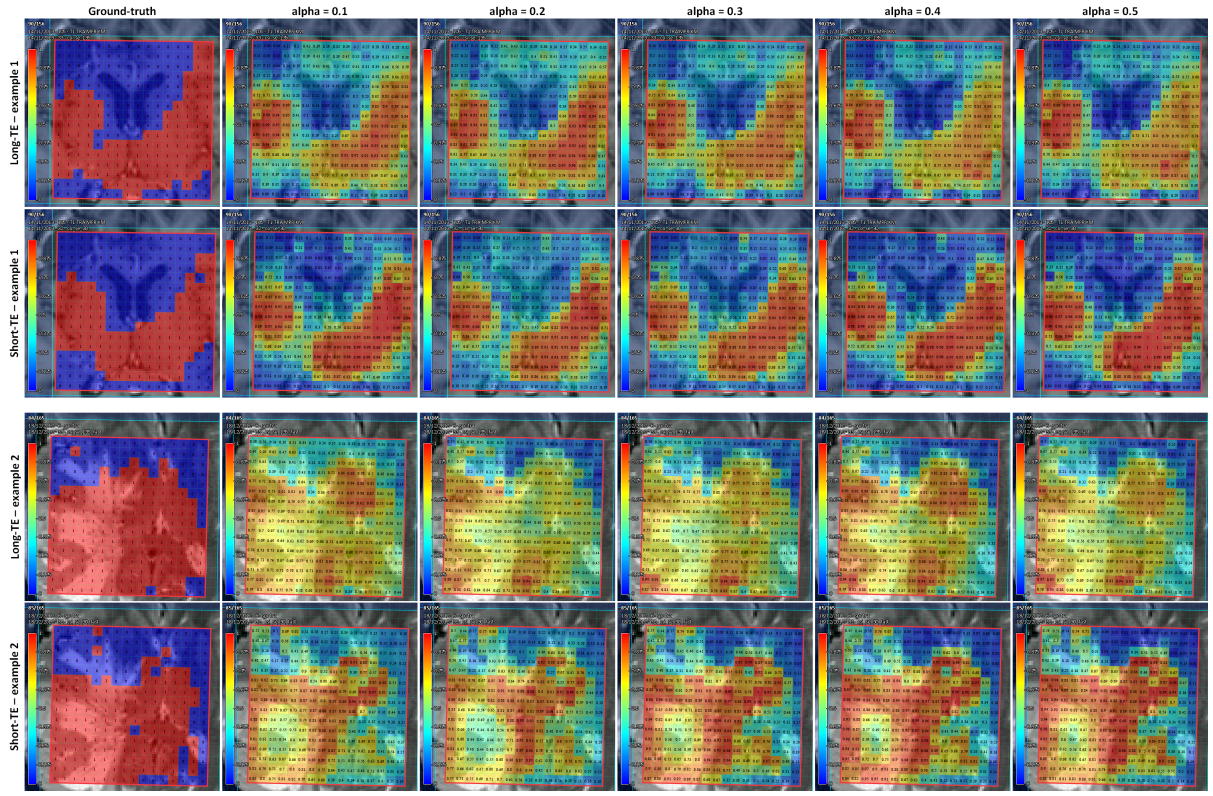

Figure 7.6: Ground-truth and classifier predictions for short and long-TE data of 2 spectroscopy studies. The predictions are shown for classifiers trained using different α values. For each MRSI example a different classifier was used where the data from the same patient was not used for training.

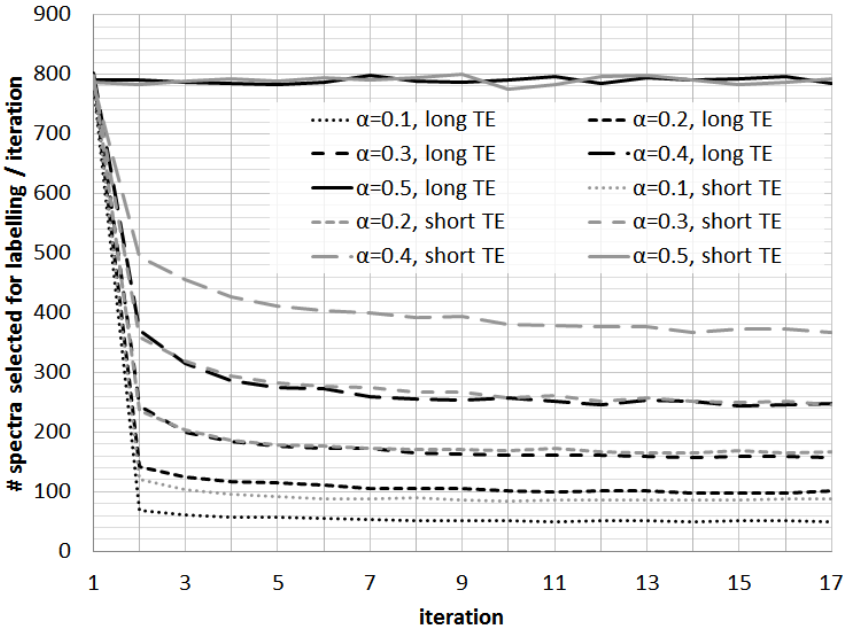

Figure 7.7: Number of spectra selected for labeling in each iteration, for both short and long TE and for different values of α. Uncertainty sampling is only performed after the first iteration, given that all spectra from the first patient are selected to train the first classifier.

7.4.4 Percentage of spectra added for each new patient

Figure 7.7 shows the number of spectra that were identified as uncertain and that were added to the training-set in each iteration. The curves confirm that the value of α allows controlling the % of spectra that a rater needs to label. Moreover, it is seen that more spectra were identified as uncertain in the first iterations, without considering the very first iteration where all spectra were used to train the first classifier. Another effect seen in this figure, is that more spectra are identified as uncertain, for the same α<0.5, in short TE than in long TE simulations. This is well matched with the performance results, which showed that proper discrimination between "accept" and "reject" of short TE data was more challenging than of long TE data.

7.4.5 Class balance

Figure 7.8 shows the effect of the proposed strategy in the class balance of the training-set. In these plots the mean % of spectra with label "accept" (henceforth %*accept*) is initially equal to the average %*accept* found in each patient (which does not consider the differences in number of spectra between patients). As more patients are added, the %*accept* converges in the case of α=0.5, to the mean %*accept* found over all spectra of the respective dataset (long or short TE). More interestingly, for all other $\alpha \neq 0.5$, the proposed approach has a gradual positive effect on the class balance as more and more data is added.

7.5 Discussion

7.5.1 Performance impact vs time-saving

The results presented above support the idea that the proposed method allows reducing the data needed to train the classifiers drastically, with minor or no impact on their performance.

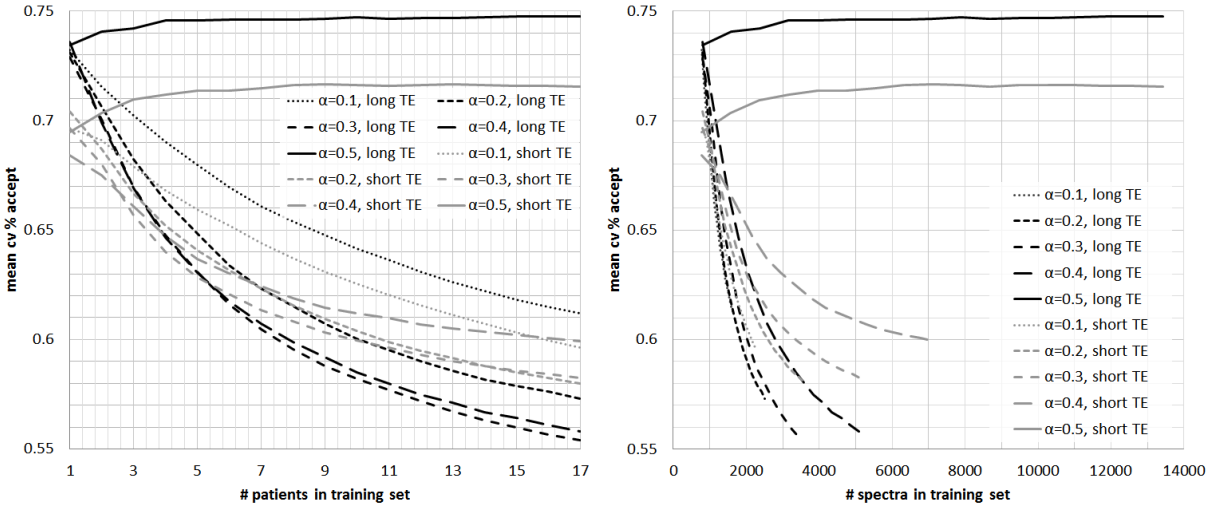

Figure 7.8: Mean % of spectra with label "accept" as a function of the number of patients (left) and as a function of the number of spectra (right) added to the training-set, for both short and long TE and for different values of α.

This increase in labeling efficiency can be seen in three different perspectives:

1. If the amount of data and the time available are limited, a rater can use the proposed approach to speed-up the labeling process. In this case the choice of the α value will reflect the personal relative importance between labeling time and performance reduction.

2. If there is a large amount of data available, the best classifier will be obtained, labeling the maximum number of different patients/cases. A small value of α is therefore recommended, as this corresponds to the most efficient way of labeling the data.

3. If there is not a lot of data but there are no time restrictions, a rater might either label the dataset completely, or choose a value of α with no significant impact on the performance.

Another aspect that should be taken in consideration is the difference between the impact in AUC and in accuracy. AUC is more sensitive than accuracy to changes in α. This can be explained by the fact that accuracy is only affected by wrong class predictions, being insensitive to variations in the outputted class-probabilities that do not change the attributed label. On the other hand, AUC measures the probability that for a randomly selected pair containing one "accept" and one "reject" examples, the %trees that vote for the class "accept" is higher for the "accept" example. Consequently, small variations in class probabilities that do not change the attributed label, can have an impact in AUC.

7.5.2 Short vs long-TE

The choice of the adequate α might be different depending on the TE. The same value of α led to the selection of more training data using short-TE than using long-TE data. Moreover, the relation between α and the impact on performance was different for each dataset.

7.5.3 Class balance

Regarding the positive effect that the proposed strategy had on the class balance, this might be connected to the relation between the training-set size and the uncertainty of the predictions seen in the results of Figures 7.6 and 7.7. As it was seen, when the training-set contains less

examples, the classifier produces more uncertain predictions. Therefore, a possible interpretation of the results of Figure 7.8 might be that classes with smaller number of examples have more uncertain predictions, and consequently, will be more likely selected with uncertainty sampling. The higher probability of selecting examples of classes containing less members helps balancing the training-set.

7.6 Conclusion

The results show that the proposed labeling strategy allows a considerable increase in the labeling efficiency. This can be translated into less labeling time needed to obtain the same classification performance, or into achieving a higher classification performance with the same labeling time. Consequently, the proposed method contributes to an easier and better implementation of methods for automatic quality control of MRSI data.

7.7 Acknowledgements

This work was funded by the EU Marie Curie FP7-PEOPLE-2012-ITN project TRANSACT (PITN-GA-2012-316679) and the Swiss National Science Foundation (project number 140958).

Part III

Assisting the interpretation of brain tumor MRS data

8 | Highlighting differences between GBM and metastasis using cNMF

This chapter is based on the following material:

Pedrosa de Barros, NM, Mocioiu, V, Ortega Martorell, S, Knecht, U, Arús, C, Slotboom, J and Julià-Sapé, M.,*Highlighting differences between GBM and brain metastasis using a blind source separation method applied to MRSI data*, ESMRMB 2015, Edinburgh, UK

8.1 Purpose

The differentiation between GBM and Metastasis is one of the most challenging tasks in Neuroradiology. Here we describe exploratory work on MR spectroscopic imaging (MRSI) data from both metastasis and GBM patients using Convex non-Negative Matrix Factorization (cNMF), a blind source separation method.

8.2 Data

A total of 38 preoperative fast-MRSI studies (1.5T, PRESS, water-suppressed, TE 135ms) from patients with GBM (n=32) and brain metastasis (n=6) were analyzed retrospectively. The 32x32 (interpolated from 12x12) MRSI grids contained spectra from both tumor and healthy tissue. Only spectra contained in the PRESS box (excluding the border ones for their lower SNR) were included, leading to a total of 7406 analyzed spectra.

8.3 Convex non-Negative Matrix Factorization

Convex non-Negative Matrix Factorization (cNMF) [Ortega-Martorell 2012a, Ortega-Martorell 2012b] is a method that allows describing a group of data-vectors - in this case spectra - as a positive linear combination of K sources, with K being user-defined (Figure 8.1). One of the main advantages of cNMF is the interpretability of its results, since the sources generated using this method are a convex combination of the data. Consequently, when applied to a group of spectra, the sources generated using cNMF can be interpreted as real spectra.

The columns of the mixing matrix **G** (reconstruction vectors) can be normalized to 1, so that its values may be interpreted as the proportion of each source used in the reconstruction of each spectrum of the analyzed dataset.

In this work cNMF was applied with $K = 3$ and **G** was normalized as described.

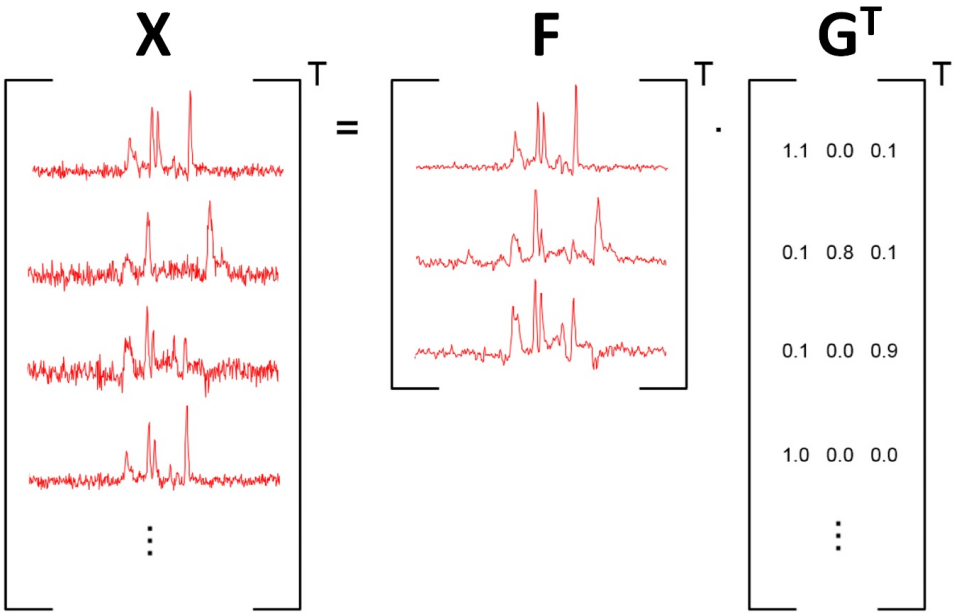

Figure 8.1: Example showing cNMF (with K=3) applied to a dataset (**X**) containing several spectra. The matrix **F** contains the generated sources, and the matrix **G** (mixing-matrix) the factors for reconstructing the data as a linear combination of sources.

8.4 Nosologic Maps

Nosologic maps were produced for each MRSI grid using SpectrIm. In those maps the values of Red, Green and Blue are given by the ratios found in the normalized mixing matrix for each source (Figure 8.2). In this way, different spectra, with different reconstruction vectors, are mapped using different colors.

8.5 GBM vs Metastasis

The differences between the MRSI maps obtained for the GBM and for the metastases were analyzed using the t-test statistic, comparing both the mean and standard deviation of the values found in the normalized mixing matrix for each source, in each MRSI grid.

8.6 Results

The obtained sources, and nosologic images for each MRSI grid are presented, respectively, in Figures 8.3 and 8.4. Figure 8.5 compares both GBM and Metastasis in terms of the obtained normalized-G values for each CSI grid. Figure 8.5 shows the comparison between GBM and metastasis in terms of the mean and standard deviation of the mixing matrix values for each source obtained for all the MRSI grids. The most significant differences were found in the "red source" maps, namely in their heterogeneity, here represented by the standard deviation. Significant differences were also found in the standard deviation of the "green" source maps.

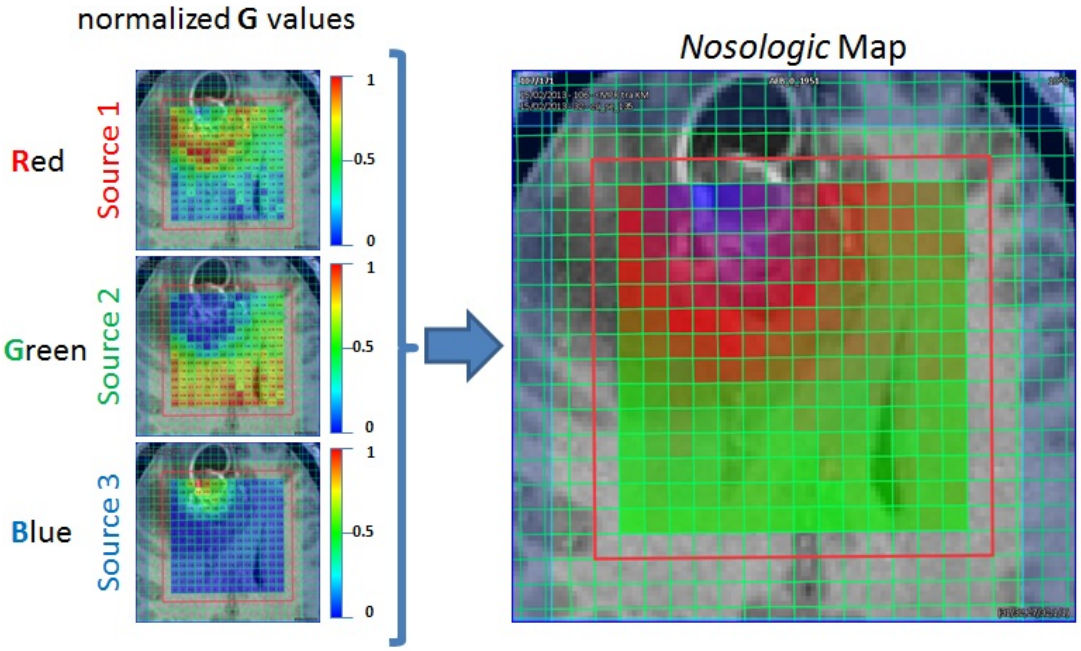

Figure 8.2: Example showing the process used to generate the nosologic maps using the normalized mixing-matrix values for each source.

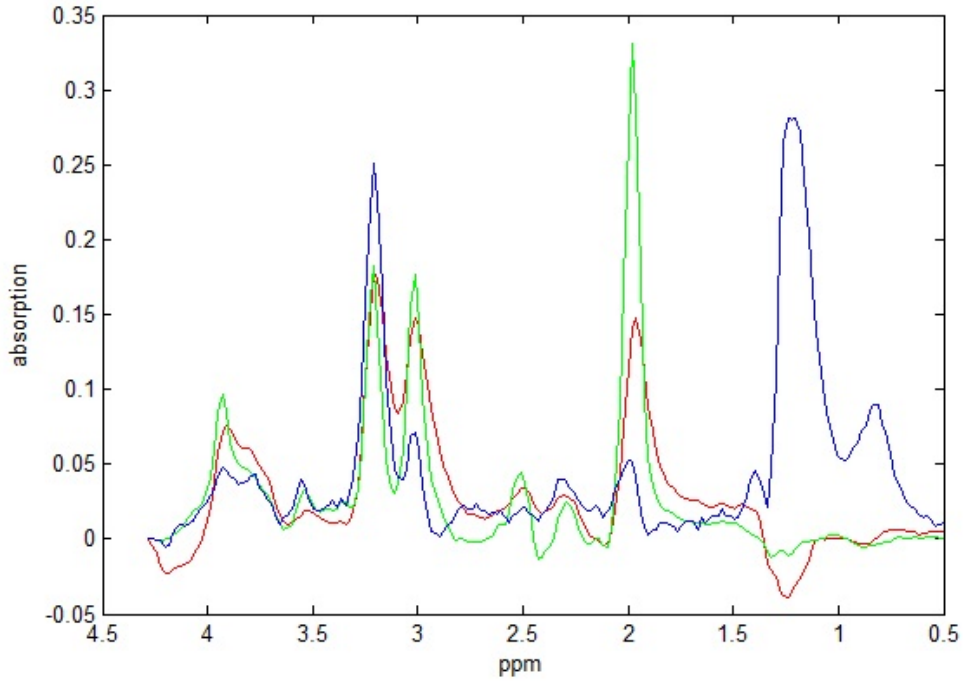

Figure 8.3: Sources obtained applying the cNMF factorization to the described GBM and Metastasis data. The "blue-source" corresponds to typical necrotic-tissue-spectra and the "green-source" to typical healthy-brain-spectra. The obtained "red-source" might be related with tumor infiltration, still, this hypothesis needs to be further evaluated.

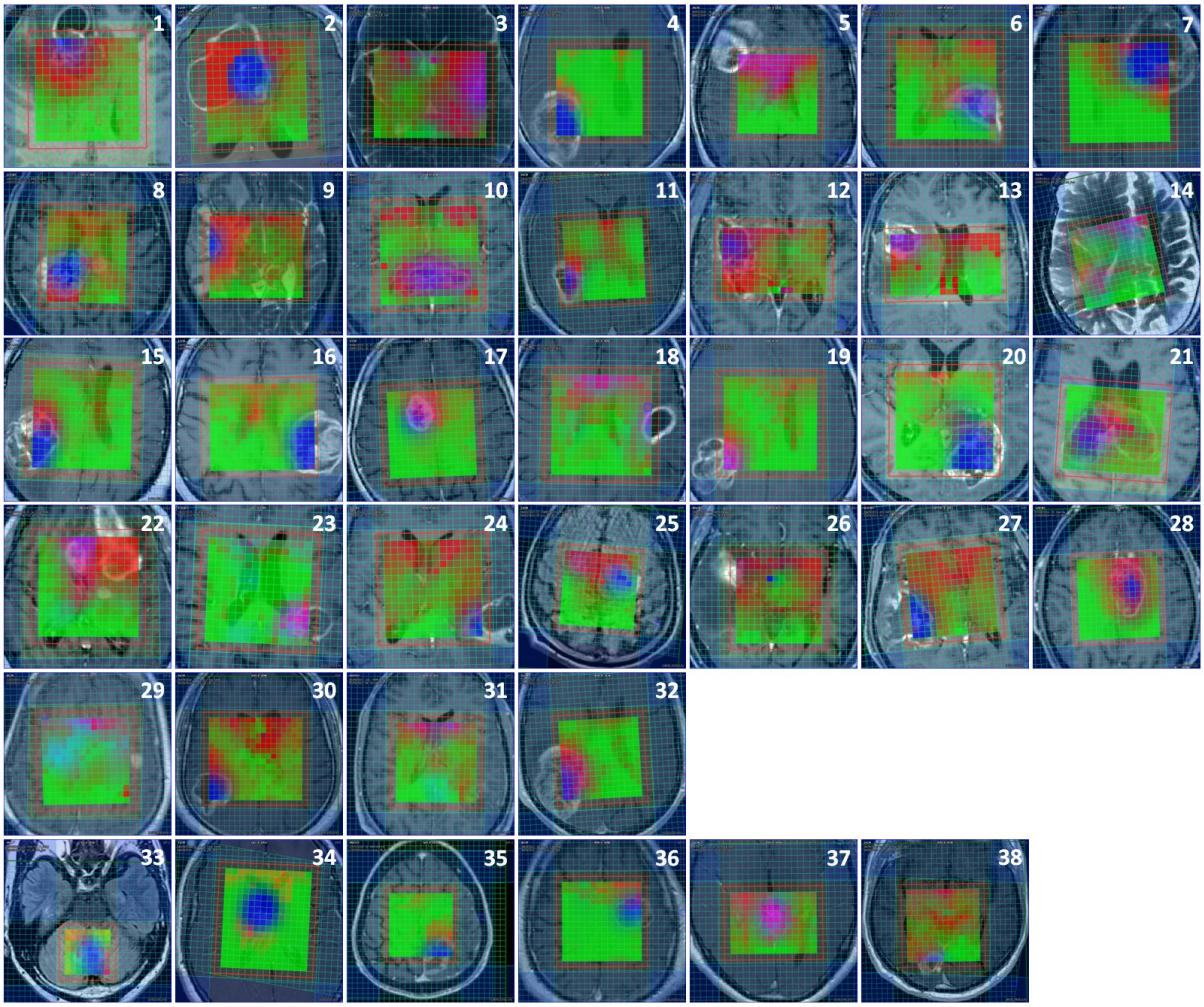

Figure 8.4: Nosologic maps obtained for the 38 analyzed MRSI grids (#1-32 GBM cases; #33-38 Metastasis cases) overlaid over T1c MRI slices (except for #14, #25, and #33 where a TRUFI sequence was used for the first one and a FLAIR sequence for the other two). The Nosologic maps were created as described in Figure 8.2 and the relation between color and source is the same as the one shown in Figure 8.3 (Green-healthy spectrum; Blue-Necrotic tumor core; Red- possibly related with infiltrated tissue but still requiring validation).

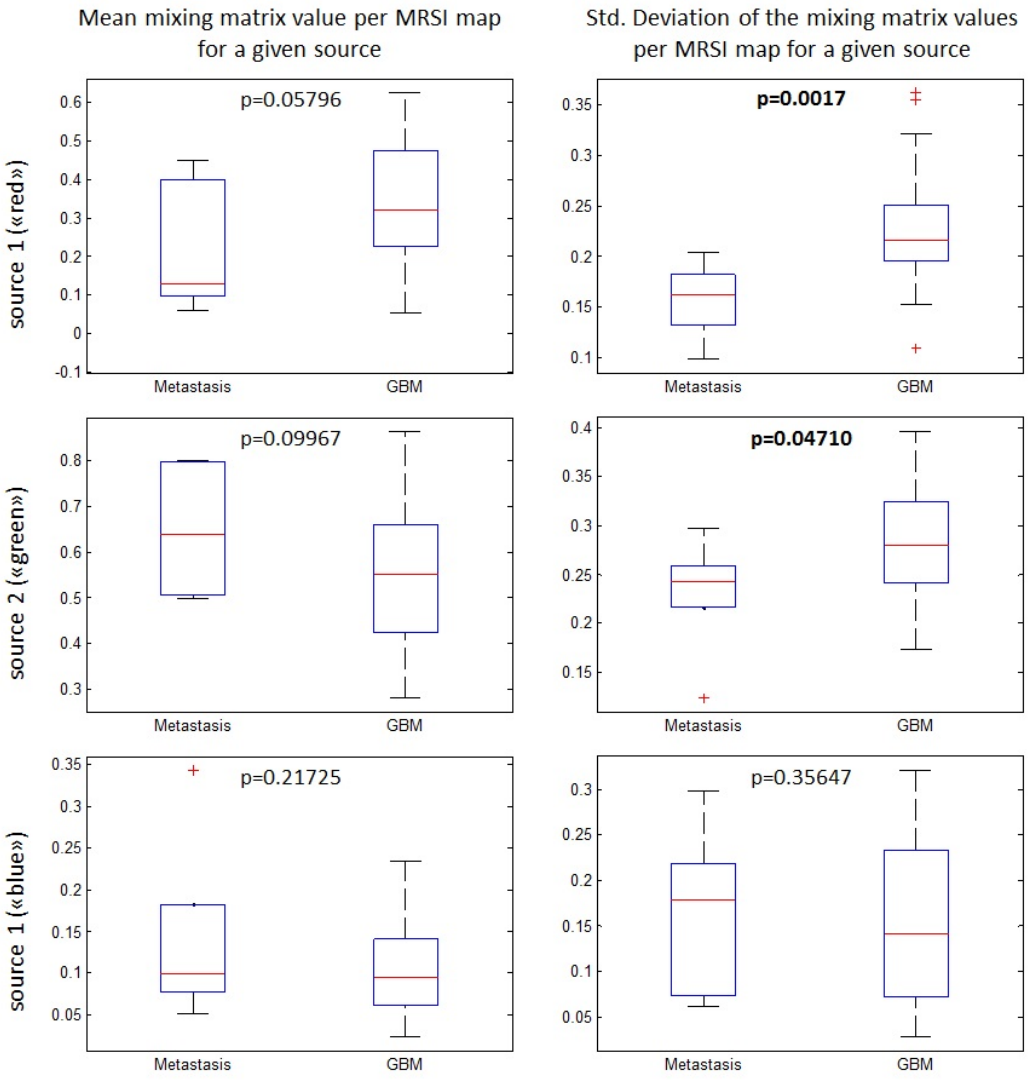

Figure 8.5: Box and whiskers plot comparing metastasis against GBM cases in terms of the mean and standard deviation of the mixing matrix values for each source obtained for all the MRSI grids. The plots show either the mean (left) or the standard deviation (right). In each comparison is included the p-value associated with the differences in the corresponding measure found between GBM and metastasis cases (t-test performed considering a single tail and 2 samples with unequal variance). The significant results ($p<0.05$) are shown in bold.

8.7 Discussion/Conclusion

The results confirm the idea that the biggest differences between GBM and Metastasis, regarding spectroscopic information, are found in the periphery of the tumor[Wijnen 2012], given that the best discriminating source ("red-source") is found frequently in this region. We hypothesize that the presence of this source might be correlated with tumor infiltration, a hypothesis which will require further evaluation. It is, in our perspective, interesting to notice that the main differences between these two groups are found in the heterogeneity of these maps, rather than in their mean values.

Despite these preliminary results, where it was possible to highlight differences between GBM and metastasis cases using cNMF, these findings do not yet allow an accurate discrimination between the two groups of tumors. Moreover, the dataset analyzed here contains significantly less metastasis examinations than GBM examinations, which might affect considerably the results. To correct this, a more balanced dataset should be used.

Another aspect to be considered is that metastases from different primary tumors have different characteristics, with certain types typically infiltrating more than others [Baumert 2006]. Consequently, instead of considering all metastases as one homogeneous group, it might be interesting to look at different types of metastases separately. This will require more cases for each metastasis type, which will also be beneficial for balancing the number of cases between GBM and metastasis.

9 | Automatic tissue-type classification of ^{1}H-MRSI spectra in patients with glioblastoma

This chapter is based on the following material:

N. Pedrosa de Barros, R. Meyer, M. Pletscher, U. Knecht, M. Reyes, R. Wiest, J. Slotboom, *Automatic tissue-type classification of ^{1}H-MRSI spectra in patients with glioblastoma*, ISMRM 2017, Honolulu, USA (**ISMRM Merit Award - Magna Cum Laude**)

9.1 Synopsis

Clinical-routine MRSI-data analysis is commonly performed through visual inspection of multiple metabolite and metabolite-ratio maps, and aims at translating the different spectroscopic patterns into known tissue-types, such as, necrosis, solid tumor, tumor-infiltration, normal-brain-tissue, etc. Such translation/segmentation requires solid expertise in MR-spectroscopy, which most clinicians do not have. Bad-quality-data, as well as non-homogeneous and frequency-dependent-selection-profiles further complicate proper interpretation of MRSI-data. Therefore, to ease the clinical-use of MRSI, we present an automatic MRSI-tissue-type segmentation algorithm, that includes automatic-quality-filtering and selection-profile-correction. The method was tested in glioblastoma and the tissue-types were compared against an MRI-based tumor-segmentation-method.

9.2 Purpose

The work here presented had the following goals:

1. To identify the most important spectroscopic patterns characteristic of different tissue-types present in ^{1}H-MRSI of glioblastoma.
2. To develop a tissue-classification-method for segmenting MRSI maps of glioblastoma patients.
3. To compare the spatial-distributions of the different identified MRS-based tissue-types with the tissue-types identified by an MRI-based automatic-tumor-segmentation-method, BratumIA [Porz 2014, Meier 2016].

9.3 Methods

The development of the MRSI-segmentation method was made in two steps. First, the main spectroscopic patterns that are present in MRSI data of gliobastoma patients were identified by clustering a training-set containing data from 17 different patients (Figure 9.1). Second,

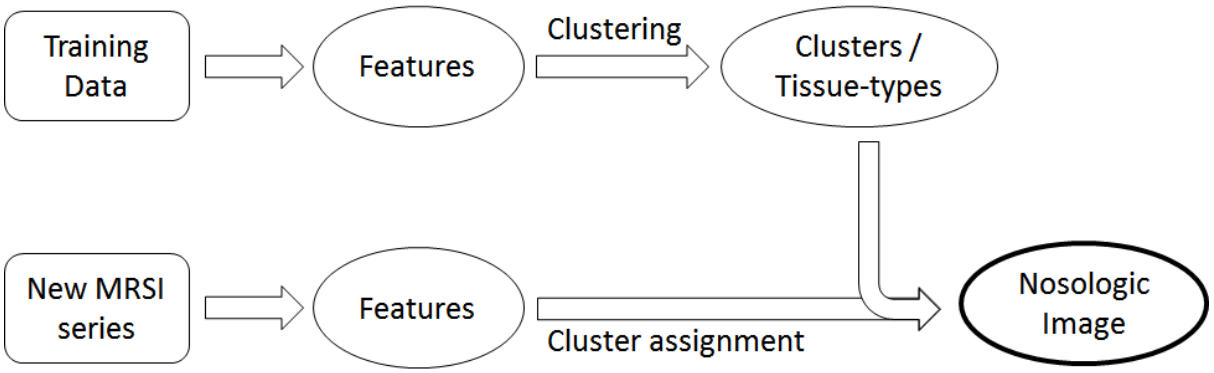

Figure 9.1: Diagram showing the pipelined used to learn the spectroscopic patterns and identify those patterns to classify new MRSI examinations.

the mean feature-values of the different identified clusters were used to classify the spectra of 5 new patients, thus segmenting the MRSI-grid into the previously-identified tissue-types. An important aspect of the strategy used here is the choice of the clustering method. For that purpose, X-means [Pelleg 2000] was used, which has the advantage that it determines the number-of-clusters based on the data, not requiring any assumption regarding the number of existent spectroscopic-patterns/clusters beforehand. Moreover, given that it works in a hierarchical fashion it allows for the detection of clusters of different sizes, what is especially relevant for this problem, given that the number of healthy-brain-tissue voxels is normally considerably larger than that of tumorous-tissue voxels.

The 22 pre-operative imaging studies of glioblastoma patients used in this study contained imaging (T1, T2, T1c, FLAIR) and ^{1}H-MRSI (2D-PRESS, CHESS, TE=135ms, TR=1500ms; 32x32 interpolated from 12x12), and were acquired on two 1.5T Siemens scanners (Aera, Avanto). The measurements were performed conform the local and national ethical regulations: all patients gave their informed consent to use the data for scientific purposes.

All spectra were pre-processed using jMRUI's SpectrIm plug-in (residual-water-removal HLSVD, automatic-quality-filtering [Pedrosa de Barros 2016b], frequency-shift-correction, auto-phasing). Quantification was performed with QUEST [Ratiney 2004], using a model containing the following metabolites: Cho, Cr, NAA, Lip1.3, Lac, Lip0.9, mIno, and Glx. In this work, Glx, mIno and Lip 0.9 were not included in the features used for clustering the data. Lactate and Lip1.3 were combined given the common errors in resolving these two peaks. The quantification results were corrected for differences in the RF-pulses' selection profile and chemical shift displacement errors, using pre-acquired MRS-phantom data.

For the clustering using X-means the following features were used: Cho/Cr, NAA/Cho, NAA/Cr, (Lip1.3+Lac)/Cr, (Lip1.3+Lac)/NAA, (Lip1.3+Lac)/Cho). For each cluster, a tissue-type/metabolic-state was assigned by an experienced spectroscopist, based on the corresponding cluster-mean feature-values and cluster-mean spectra. The MRSI-segmentation performed on the 5 test studies, assigned to each spectroscopic-voxel the closest cluster. The metabolite maps were interpolated to 64x64 prior to MRSI-segmentation and the resolution of the BratumIA segmentation was downscaled to match the resolution of the spectroscopy. MRI segmentation was performed using BratumIA [Porz 2014, Meier 2016], an automatic segmentation method that was trained on manually segmented MR-images of glioblastoma patients. In order to be applied, BratumIA requires as input T1, T1c, T2 and FLAIR, and segments the images into the following tissue-types: white-matter, grey-matter, CSF, edema, non-enhancing-tumor,

enhancing-tumor and necrosis. The results of MRSI and MRI segmentations were compared.

9.4 Results/Discussion

Figure 9.2 shows the different representative cluster spectra of the identified clusters, which were ordered according to their degree of change relatively to healthy brain spectra. The different clusters were grouped into 4 main groups: Healthy, Infiltration, Solid Tumor and Necrosis. All these associations require further validation with histopathology. In the case of the Infiltration clusters, given that this group shows important but very subtle differences with healthy brain spectra (increased Cho/NAA ratio relatively to normal brain tissue [Croteau 2001]), a question mark was added to the name of the group to indicate the uncertainty of the class.

In Figures 9.3 and 9.4 the different MRSI-segmentations are shown together with the corresponding MRI (BraTumIA) segmentations. Finally, in Figure 9.5 the confusion matrix comparing both types of segmentation is shown. The table shows the number of voxels of the test-set assigned to each cluster, as well the ratio of voxels from each cluster assigned to each of the MRI-segmentation tissue-types. Besides a general good agreement between both types of segmentation, it should be stressed that the proposed method detected patterns that are expected to be associated with tumor infiltration in regions identified by BratumIA as white-matter. These results are in agreement with previous results from [Cordova 2016] and [Stadlbauer 2004], showing that Cho/NAA abnormalities extend beyond structural imaging abnormalities and can improve brain tumor delineation.

9.5 Conclusion

A novel method for mapping different brain and brain-tumor tissue types based on 2D-MRSI was presented. The method assigns tissue types to spectral patterns, and may assist clinicians interpreting MRSI examinations of brain tumors. One of the most interesting aspects of the results presented is the detection of patterns that may be associated with tumor infiltration. Given that the association between spectra and tissue-types was done using information from several spectroscopic features, it is expected that this approach may be used to increase the specificity of the detection of patterns such as infiltration, in comparison with approaches that rely solely on a single parameter (e.g. detection of tumor infiltration using only Cho/NAA ([Cordova 2016]). This is, however, an aspect that requires further validation.

9.6 Acknowledgements

This work was funded by the EU Marie Curie FP7-PEOPLE-2012-ITN project TRANSACT (PITN-GA-2012-316679) and the Swiss National Science Foundation (project number 140958).

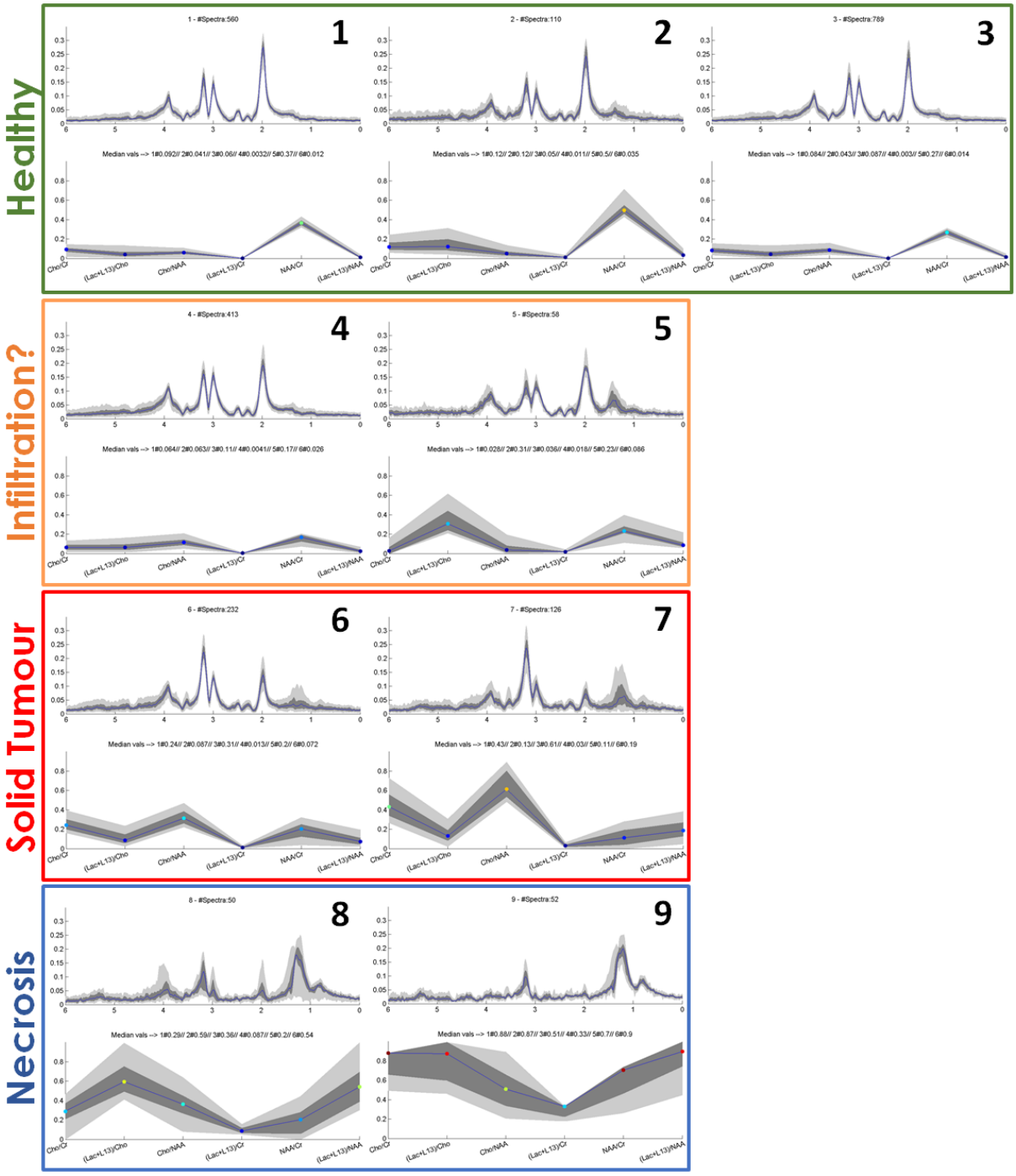

Figure 9.2: Different representative cluster spectra of the identified clusters, which were ordered according to their degree of change relatively to healthy brain spectra. For each cluster, the corresponding normalized feature vector is also shown. The normalization was done by first performing 99% winsorization followed by rescaling to $[0,1]$ interval.

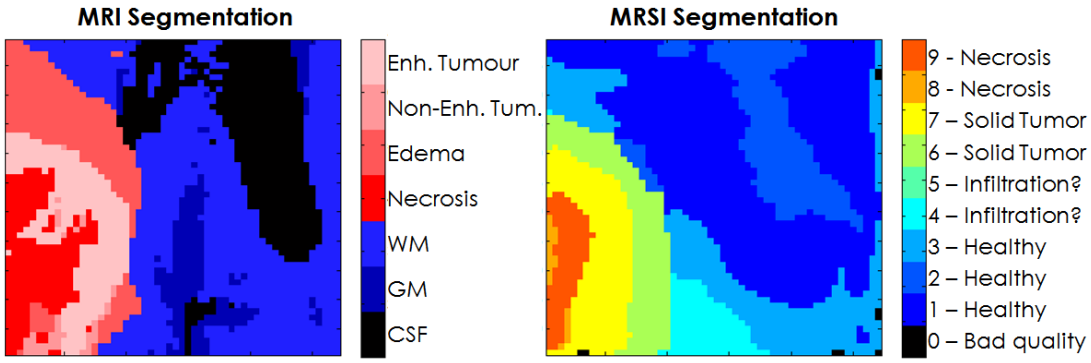

Figure 9.3: Comparison between the MRI and MRSI segmentations for one of the cases included in the test-set. MRI segmentation performed using BratumIA.

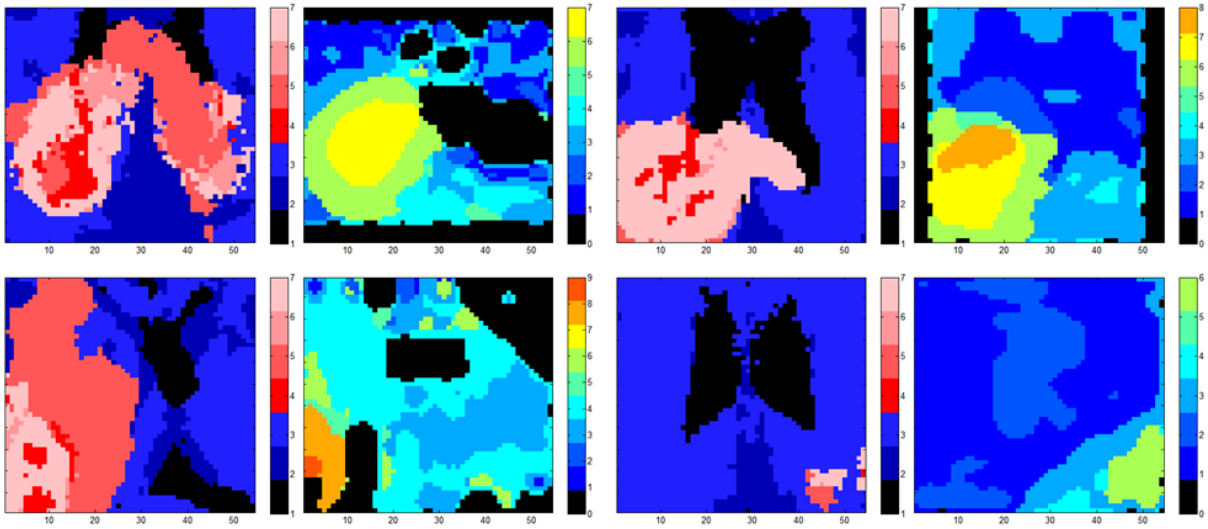

Figure 9.4: Comparison between the MRI and MRSI segmentations for 4 of the cases included in the test-set. MRI segmentation performed using BratumIA.

	#voxels	CSF	GM	WM	Edema	N.E. Tum.	E. Tum.	Necrosis
1 – Healthy	4265	27.2%	5.5%	61.6%	4.3%	0.5%	0.9%	0.0%
2 – Healthy	1389	45.4%	5.5%	42.4%	3.5%	0.4%	2.8%	0.0%
3 – Healthy	2705	7.8%	18.5%	58.1%	11.9%	0.8%	2.7%	0.1%
4 – Infiltration?	1699	11.6%	22.4%	23.8%	38.6%	1.6%	1.9%	0.1%
5 – Infiltration?	120	5.0%	25.0%	13.3%	30.0%	1.7%	24.2%	0.8%
6 – Solid Tumor	1304	3.9%	3.0%	17.9%	13.9%	6.4%	50.5%	4.4%
7 – Solid Tumor	844	0.0%	0.0%	0.2%	6.8%	6.5%	53.6%	32.9%
8 - Necrosis	304	1.6%	0.0%	0.0%	1.6%	7.2%	52.0%	37.5%
9 - Necrosis	142	0.0%	0.0%	0.0%	0.0%	1.4%	25.4%	73.2%

Figure 9.5: Table showing the number of voxels from the test-set assigned to each cluster, as well as the ratio of the voxels from each cluster associated with each of the tissue-types outputted by BratumIA. (GM=grey-matter; WM=white-matter; NE Tum.=non-enhanching tumour; E Tum.=enhancing tumour).

10 | BraTSIS: Brain Tumour Spectroscopy Imaging Segmentation protocol

Authors: Nuno Pedrosa de Barros, Raphael Meier, Martin Pletscher, Samuel Stettler, Urspeter Knecht, Mauricio Reyes, Jan Gralla, Roland Wiest, Johannes Slotboom.

The content of this chapter was *submitted* for publication to the journal *NMR in Biomedicine*.

10.1 Abstract summary

Clinical-routine MRSI-data analysis is commonly performed through the visual inspection of multiple metabolite and metabolite-ratio maps and aims at *translating* the different spectroscopic patterns into known tissue-types, such as necrosis, solid tumor, tumor-infiltration, normal-brain-tissue, etc. Such *translation/segmentation* requires solid expertise in MR-spectroscopy, which clinicians often do not have. Therefore, to ease the *clinical-use* of MRSI, we present a new method named BraTSIS (Brain Tumor Spectroscopic Imaging Segmentation) that allows identifying the main spectroscopic patterns seen in GBM patients. The method was developed using 50 MRSI examinations (2D PRESS, TE 135ms) of 30 GBM patients and tested in an extra 20 MRSI examinations from 12 other patients. MRSI data was processed using jMRUI's SpectrIm plugin and included automatic-quality-control and selection-profile-correction. The MRI-segmentation performed by BraTumIA, which is an automatic image segmentation method, was used for the development and validation of BraTSIS. The comparison between BraTSIS and BraTumIA provided further insights into the complementary aspects of MRSI and structural MRI. The BraTSIS protocol aims at supporting the clinical interpretation of malignant glioma MRSI data and represents an effort towards the standardization of MRSI analysis, which is essential for the further development of MRSI as a reference technique for the *non-invasive* characterization of brain tumors.

10.2 Introduction

Magnetic Resonance Spectroscopy (MRS) provides relevant information for the assessment of brain tumors, allowing to distinguish different tumor types and grades [Zou 2011, Zeng 2011, Law 2003], distinguish radiation effects (pseudoprogression) from true progression [Matsusue 2010, Lichy 2004] and identify regions with high tumor cellularity that are not visible in structural MRI (sMRI) [Cordova 2016, Guo 2012, Stadlbauer 2004]. Many publications [Hollingworth 2006] focus on the translation of one or two MRS features, such as Cho/NAA and Cho/Cr, into clinically meaningful information for the tasks described above. Regardless of what can be achieved with *single-feature-MRS-analysis*, multi-feature-analysis of MR-spectra has the potential to allow better characterization of the different tissues in brain tumor patients.

However, the combined analysis and the interpretation of the various metabolites and metabolite ratios is complex[García-Figueiras 2016] and requires solid expertise in MRS. Furthermore, there is a lack of consistent guidelines for the use of the information provided by the different MRS features[Oz 2014], which complicates the identification of consensus on the clinical meaning of the data obtained using MRS. Therefore, MRS interpretation varies depending on the place and person who performs the analysis.

To facilitate the clinical interpretation of MR spectra, different authors [Preul 1996, Preul 1998, Luts 2009, Li 2013a, Raschke 2015, Yang 2015] have proposed various methods that support the translation of spectra into different tissue-types. These methods require a library of spectral patterns, where a tissue-type is assigned to each pattern. Once the translation is performed, the results can then be presented in the form of *nosologic* images, images where color is used to identify different tissue-types or diseases. The dimensionality reduction together with the improved data-visualization provided by these methods reduces the complexity of the analysis and facilitates the clinical interpretation of MRSI data. Despite the very interesting results, the translation of the proposed approaches into a standard for MRSI data analysis is complicated, requiring the libraries of the considered spectral patterns to be shared between centers. Moreover, the list of GBM-specific patterns considered in previous libraries is not complete, missing, for instance, patterns with high levels of Glx.

To tackle these problems, we performed a thorough analysis of the main types of spectra existent in 50 MRSI examinations (2D PRESS, TE 135ms) of 30 GBM patients, and developed a simple decision-tree based protocol to recognize the identified tissue types. The development and evaluation of the BraTSIS (Brain Tumor Spectroscopic Imaging Segmentation) protocol were supported by BraTumIA (Brain Tumor Image Analysis)[Meier 2016, Porz 2014], which is an automatic brain tumor segmentation method trained to reproduce the manual MRI segmentation performed by experienced neuroradiologists. The BraTSIS protocol aims at supporting the clinical interpretation of MRSI data of patients with malignant glioma and represents an effort towards the standardization of MRSI analysis, which is essential for the further development of MRSI as a reference technique for the *non-invasive* characterization of brain tumors.

This paper is organized as follows. First, the spectra of the training dataset were clustered using X-means to help to visualize the different spectroscopic patterns existent in MRSI data of GBM patients. The representative spectra of the differently identified clusters were essential for defining the different classes that were considered in BraTSIS, as well as the structure of the decision tree that was developed to identify them. For the definition of the decision rules used in BraTSIS to detect Necrosis and Solid Tumor, MRI segmentation was used as a reference. The differentiation of normal versus abnormal values for certain MRS features was made using the data collected from healthy volunteers. The paper finishes with a comparison between the segmentation results obtained using BraTSIS and the MRI segmentation performed by BraTumIA. This was performed using a test dataset containing 20 MRSI examinations and MRI segmentations of extra 12 patients that were excluded from the development phase. The comparison between the MRI and MRS segmentations provided further insights into the complementary aspects of MRSI and sMRI.

10.3 Methods

10.3.1 Data

Data acquisition was performed using two 1.5 T Siemens scanners (models 'Aera' and 'Avanto', Siemens, Erlangen, Germany). A total of 70 pre- and post-operative MRSI examinations (PRESS, CHESS water suppression, TE=135 ms, TR=1500 ms) were collected from 42 different GBM patients. The MRSI data had an original resolution of 12x12, with a voxel size of 13.33x13.33x15 mm, and were interpolated before DICOM storage to a resolution of 32x32 and a voxel size of 5x5x15 mm. All spectra from outside the PRESS-box, as well as one single outer voxel row of the excited PRESS volumes, were discarded given that often unreliable quantification results are found on the edges of the PRESS box due to partial excitation. The data was separated into a training dataset, containing 50 MRSI studies from 30 different patients and test dataset with the remaining 20 cases of an extra 12 patients. After exclusion of the outer voxels, each data set contained respectively 8604 and 3808 spectra. The test dataset was not considered during the development of the protocol and was only included for the final comparison between BraTSIS and BraTumIA segmentations. Besides spectroscopy, the protocol included acquisition of T1-weighted (MP RAGE, TE 2.67 ms; TR 1580 ms, TI 900 ms; 1x1x1 mm), T1-weighted gadolinium-enhanced (MP RAGE, TE 4.57 ms; TR 2070 ms, TI 1100 ms; 1x1x1 mm), T2-weighted (SPACE, TE 380ms, TR 3000ms; 1x1x1 mm), and FLAIR-weighted MRI (TE 88 ms, TR 8000 ms, TI 2500 ms; 1x1x3.3 mm).

In parallel to these patients, 7 MRSI normal control recordings were collected from 3 healthy volunteers using the same acquisition parameters. This healthy control dataset contained a total of 1372 spectra after exclusion of the outer-voxel row of the excited volume.

The measurements were performed in conformance with local and national ethical regulations, meaning that all patients gave their written informed consent for the data to be used for scientific purposes.

10.3.2 MRS Data Processing Pipeline

All spectra were processed using jMRUI's SpectrIm plugin (`www.jmrui.eu`). The processing pipeline included the following six processing steps:

1. Residual water peak removal using HLSVD [den Boogaart 1994];
2. Automatic Quality Control [Pedrosa de Barros 2016b, Pedrosa de Barros 2016a];
3. Frequency-Shift Correction;
4. Auto-phasing;
5. Quantification using QUEST[Ratiney 2004];
6. Selection profile correction[Pedrosa de Barros 2017b].

In order to get meaningful and consistent results, strict MRSI-quality control is essential. Therefore, all low-quality voxels (i.e. voxels with a probability of acceptance by an expert < 0.5) were excluded from the analysis. For more details on automatic quality filtering the reader is referred to [Pedrosa de Barros 2016b, Pedrosa de Barros 2016a, Pedrosa de Barros 2017b].

The metabolite model used for quantification using QUEST included the following metabolites: choline (Cho), creatine (Cr), N-acetyl aspartate (NAA), glutamate (Glu), glutamine (Gln), lactate (Lac), lipids 1.3 (Lip1.3), lipids 0.9 (Lip0.9), myo-inositol, glycine. The latter two metabolites were not analyzed in this paper given their quantification results were found to be non-reliable for the given echo time.

Excitation profile imperfections of amplitude modulated RF-pulses, which are used in the PRESS sequence, cause non-uniform spatial excitation. Additionally, the scanners' limited gradient and RF-power cause spins resonating at different chemical shifts to be excited in a spatially shifted manner, which is the so-called chemical shift artifact (CSA). These effects can be corrected using phantom MRSI recordings performed on metabolite solutions that use the same acquisition parameters as the *in-vivo* measurements. To the best of the authors' knowledge, the CSA correction applied in this study has never been used before in the study of malignant glioma brain tumors. If not corrected, this has a disastrous effect not only on the observed metabolite values but also metabolite ratios. The CSA effect is generally observed in volume localization sequences such as PRESS and STEAM, and to less extent in (semi)LASER[Slotboom 1991, Slotboom 1994b], namely for low selection RF pulse bandwidth, weak B_0 gradients and high field strength. Metabolite ratios such as Cho/NAA are considerably more affected than Cho/Cr, given that NAA has a higher chemical shift difference with Cho than Cr. Detailed information on the applied CSA and selection profile correction is found in [Pedrosa de Barros 2017b]. An average Cho/NAA map before and after correction for both healthy controls and GBM patients was included in the results section to highlight the importance of this correction.

10.3.3 Image Segmentation using BraTumIA

For each MRSI examination, the corresponding MR images were segmented into the following compartments: white-matter (WM), gray-matter (GM), cerebral-spinal-fluid (CSF), edema, non-enhancing tumor (NE-Tumor), enhancing tumor (E-Tumor) and necrosis. *MRI-visible-tumor* was defined as the union of NE-Tumor, E-Tumor, and necrosis.

The enhancing tumor is visible on T1c-weighted MR images as a hyperintense region caused by the contrast-enhancement due to the blood-brain barrier disruption. Enhancing tumor is usually surrounded by a large edema which appears hyperintense on T2-weighted images and hypointense on T1-weighted images. The FLAIR image is utilized to differentiate edema from CSF, which appears also hyperintense in the T2 sequence but dark in the FLAIR image due to the suppression of the free water signal. Glioblastoma can contain necrotic parts which do not enhance on T1c-weighted images but appear hyperintense on T2-weighted images and hypointense on T1-weighted images. The most difficult compartment to identify is the non-enhancing tumor which does not enhance in T1c-weighted images and exhibits generally a lower intensity in T2-weighted images than necrosis with the corresponding T1-hypointensity.

The segmentation of the different compartments was performed using BraTumIA [Meier 2016, Porz 2014], an automatic method that was trained to reproduce the manual image segmentation performed by experienced neuroradiologists. This method uses as input the previously introduced MR image sequences: T1, T1c, T2, and FLAIR. BraTumIA is a machine learning-based segmentation method, which was trained on an independent dataset of 54 pre- and postoperative MRI examinations of glioblastoma. It uses supervised learning to generalize from manually segmented ground truth data. A more detailed description of the algorithm and training data can be found in the study of Meier et al. [Meier 2016].

The process of manually segmenting glioblastoma includes first the segmentation of the

complete tumor visible as a hyperintense area in the FLAIR image. Second, the gross tumor volume (including necrosis, non-enhancing and enhancing tumor) is delineated using mainly the co-registered T1c and T2 sequences. This region exhibits a lower intensity in T2 than the surrounding edema. The mismatch between both regions is considered as the edema. The enhancing tumor is segmented based on the hyperintense region in the T1c sequence. Necrosis is differentiated from non-enhancing tumor based mainly on information from the T2-weighted sequences.

Image segmentation could not be performed in 4 out of the 50 training cases, given that not all required image sequences were available. To facilitate comparison with MRS, the image segmentation results were converted to the same orientation and resolution as MRSI.

10.3.4 Clustering

The spectra from the training set were clustered using the *Weka* [Hall 2009] implementation of X-means [Pelleg 2000], a hierarchical clustering method based on K-means that automatically identifies the number of clusters in a dataset. The clustering was performed using the following 17 MRS features: %NAA, %Cho, %Cr, %Glx, %Lac, %Lip, Lac/Cr, Lip/Cr, Cho/Cr, Cho/NAA, Glx/NAA, Glx/Cr, %(Lip+Lac), Lac/NAA, Lip/NAA, (Lac+Lip)/Cr, and (Lac+Lip)/NAA. Glu and Gln were not included separately given that, due to their overlap, the quantification of these metabolites on a spectrum by spectrum basis is difficult and may lead to the creation of extra non-meaningful clusters.

Prior to clustering, 99% winsorization [Tukey 1962] was applied to each feature followed by feature rescaling to [0, 1]:

$$x_i' = \frac{x_i - \min(X_i)}{\max(X_i) - \min(X_i)} \tag{10.1}$$

Feature standardization was not performed given that several of the features presented highly skewed distributions. The distance metric chosen for clustering was the Manhattan distance (L_1), which is expected to provide a better overall discrimination between the different feature vectors than the Euclidean distance (L_2) for the number of features used [Aggarwal 2001].

To test the stability of the clusters found with X-means, the training set was partitioned into 5 different groups containing 10 MRSI examinations each. The data was organized such that the examinations of each patient were all in the same group. The clustering was then performed 5 times, excluding every time one of the five partitions completely.

10.3.5 BraTSIS: MRSI Segmentation Protocol

The results of X-means showed 5 main types of spectra that are well represented in the example shown in Figure 10.1.

BraTSIS was designed to standardize the identification of these patterns considering the most relevant and state-of-the-art decision rules[Vigneron 2001, Horská 2010, Ricci 2007, Zou 2011, Zeng 2011, Law 2003, Matsusue 2010, Lichy 2004, Cordova 2016, Guo 2012, Stadlbauer 2004, Hollingworth 2006] for the interpretation of brain tumor MR spectra. Its development was driven by the following objectives:

1. The protocol should focus on the main aspects that make MRS complementary to MRI in the characterization of brain tumors, namely:

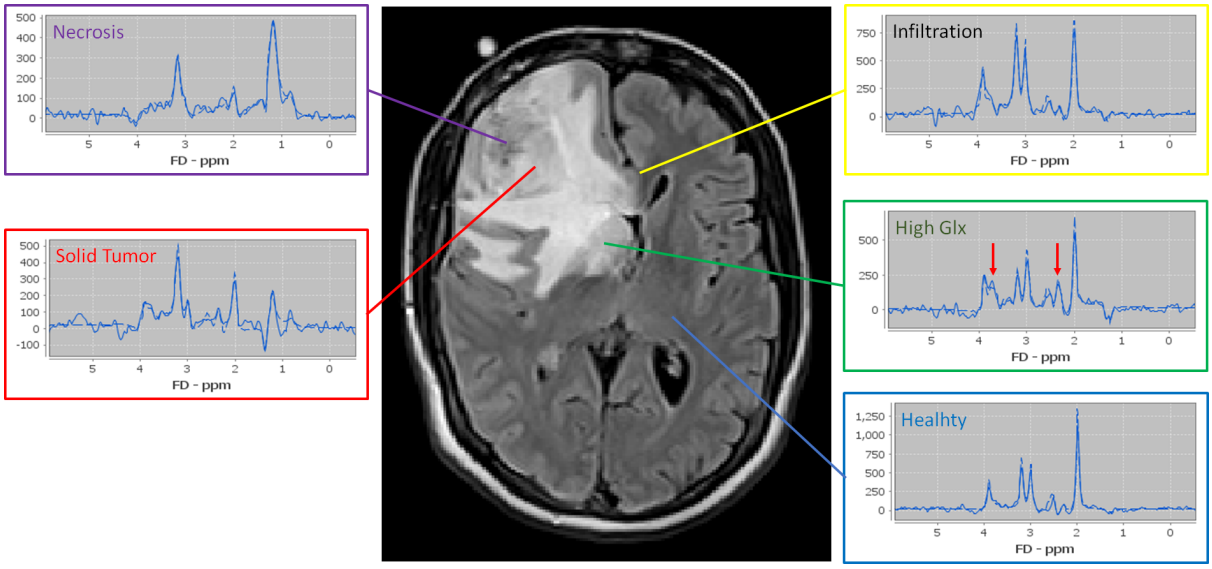

Figure 10.1: Pre-operative MRSI acquisition of a GBM patient showing all of the main types of spectra identified after clustering the training dataset using X-means, and their location. A FLAIR image was used to indicate their locations. The selected case corresponds to one of the cases of the training set.

(a) Accurate detection of tumor necrosis/cysts whose spectroscopy patterns are easily identifiable: very high levels of Lip and/or Lac and low levels of other metabolites [Vigneron 2001, Horská 2010].

(b) Detection of spectroscopy features characteristic of high-grade tumors [Zeng 2011, Law 2003], namely very high *Cho/Cr* and *Cho/NAA*.

(c) Detection of tumor infiltration [Cordova 2016, Guo 2012, Stadlbauer 2004] not visible using sMRI, allowing to improve delineation of tumor borders: *Cho/NAA* higher than in healthy brain tissue.

(d) Detection of features associated with neuronal loss and demyelination seen often in peritumoral edema of GBM patients [Ricci 2007]: *Glx/Cr* higher than in normal brain tissue.

2. The protocol should be based on simple decision rules, allowing it to be easily applied in different centers, and easily updated and adapted to different sequences and acquisition parameters, if needed.

Considering these objectives, BraTSIS was defined as a decision tree that starts by identifying the tissue types that have the biggest differences with normal brain tissue. Consequently, the first tissue-type to be detected is Necrosis, which can be easily separated from the remaining spectra by the very high levels of *Lip* and/or *Lac* relatively to the remaining metabolites. Once Necrosis spectra are excluded, the high values of *Cho/Cr* and/or *Cho/NAA* allow to identify the spectra characteristic of *Solid Tumor*. *Infiltration* is then defined as spectra which have abnormal values of *Cho/NAA* and/or *Cho/Cr* that were not included in the *Solid Tumor* class. Finally, if none of the previous tests is positive, the levels of *Glx* are evaluated. If abnormal levels of *Glx* are detected, the spectra are included in the *High Glx* group, otherwise, the spectra are considered *Healthy*. The description of each class and the strategy used to define the corresponding class membership tests are described below:

1. *Necrosis:* Spectra with very high levels of *Lip* and/or *Lac* relatively to other metabolites. Multiple membership tests using different features that capture this property of necrosis were tested and optimized to maximize the agreement (Dice-score) with the *Necrosis* class in MRI segmentation. The specific tests that were evaluated to optimally detect necrosis are listed in the next section.

2. *Solid Tumor*: Spectra with very high *Cho/Cr* and/or *Cho/NAA* associated with the *MRI-visible tumor*. Different membership tests using different combinations of these parameters were tested and optimized to maximize the agreement (Dice-score) with the *MRI-visible tumor*. The specific tests evaluated to detect *Solid Tumor* are described in the next section.

3. *Infiltration*: Spectra whose values of *Cho/NAA* and/or *Cho/Cr* are significantly higher than in normal brain tissue but lower than in Solid Tumor. Due to the importance of this class and the broad range of spectra that it contains, this class was divided into two subgroups: *Infiltration I* and *II*, the second one including the spectra with higher levels of *Cho/NAA* and/or *Cho/Cr*. The division was made simply using the mid-values between the thresholds used to detect *Infiltration I* and the ones used to detect *Solid Tumor*. The detection of *Infiltration I* is made using the 99.9^{th} percentile of *Cho/NAA* in healthy volunteers as reference.

4. *High Glx*: Spectra with abnormally high levels of *Glx* but with normal *Cho/NAA*. Given that the *Glx* increase has been associated with neuronal loss and demyelination [Ricci 2007], it is expected that the ratio with *NAA* increases the sensitivity to detect the deterioration and/or loss of neurons. Due to this reason, *Glx/NAA* was selected for the class membership test associated with the class *High Glx*. As in the detection of *Infiltration I*, the class membership test used as a threshold the 99.9^{th} percentile of the *Glx/NAA* in healthy volunteers.

5. *Healthy*: Applies if none of the previous class membership conditions is met.

For the classes *Infiltration I* and *High Glx*, where the threshold values were determined as the 99.9^{th} percentile of the selected features in healthy volunteers, the associated 95% confidence intervals were calculated using the percentile method [Efron 1981]:

$$Prob\left\{p_{99.9} \in \left[x_{i,(k_1)}, x_{i,(k_2)}\right]\right\} = \sum_{k=k_1}^{k_2-1} b_{k,n}(0.999) \tag{10.2}$$

, where $b_{k,n}(p)$ is the binomial distribution:

$$b_{k,n}(p) = \left(\begin{array}{c} n \\ k \end{array} \right) p^k (1-p)^{n-k} \tag{10.3}$$

, $x_{i,(1)} < x_{i,(2)} < \cdots < x_{i,(n)}$are the order statistics for the feature i, n denotes the size of the sample collected from healthy volunteers, and k is the position of a given element of the order statistics of the analyzed feature. Prior to determining the percentiles and corresponding confidence intervals, the size of the healthy control dataset was artificially increased interpolating each feature map by a factor of 4 (zero-filling) in each spatial direction.

The different features and thresholds selected to be used in BraTSIS are presented in the results section (Figure 10.6 and Table 10.1).

The structure and classes of BraTSIS were determined based on the representative spectra and features values of the clusters identified by X-means. To ensure that the rules defined to classify the clusters identified by X-means applied to the classification of individual spectra, the following test was performed. First, the different clusters obtained for each of the five different solutions of X-means were classified using BraTSIS. The clusters of the training set were then used to classify the spectra of the test set. This was done by assigning to every test-set spectrum the class of the closest training set cluster, i.e. the cluster whose median feature values had the smallest L_1 difference with the feature values of the new spectrum. These are named as the *indirect* BraTSIS results, which varied depending on the clustering solution used. The *direct* BraTSIS results were obtained simply by classifying each new spectrum directly. Both approaches were compared by measuring the average agreement between the direct BraTSIS and the different indirect BraTSIS results, and comparing it to the average agreement between the indirect BraTSIS solutions. The agreement was evaluated as the Dice-Score for each class. The results (included in part B of the Supplementary Material) showed that both approaches produce nearly identical results, confirming the applicability of the rules to the classification of individual spectra.

10.3.6 Optimized detection of Necrosis and Solid Tumor

The decision rules used in BraTSIS to assign new spectra to the *Necrosis* and *Solid Tumor* classes were optimized using MRI Segmentation as a reference. For the necrosis class, several decision rules were tested that aim at detecting *high* levels of Lip and/or Lac and *low* levels of other metabolites:

1. $(Lac+Lip)/NAA \geq \theta_{a.1}$
2. $(Lac+Lip)/Cr \geq \theta_{b.1}$
3. $Lac/Cr \geq \theta_{c.1}$
4. $Lip/Cr \geq \theta_{d.1}$
5. $Lip/NAA \geq \theta_{e.1}$
6. $Lac/NAA \geq \theta_{f.1}$
7. $(Lac+Lip)/NAA \geq \theta_{g.1}$ *AND* $(Lac+Lip)/Cr \geq \theta_{g.2}$
8. $(Lac+Lip)/NAA \geq \theta_{h.1}$ *OR* $(Lac+Lip)/Cr \geq \theta_{h.2}$
9. $Lac/Cr \geq \theta_{i.1}$ *AND* $Lip/Cr \geq \theta_{i.2}$
10. $Lac/Cr \geq \theta_{j.1}$ *OR* $Lip/Cr \geq \theta_{j.2}$
11. $Lac/NAA \geq \theta_{k.1}$ *AND* $Lip/NAA \geq \theta_{k.2}$
12. $Lac/NAA \geq \theta_{l.1}$ *OR* $Lip/NAA \geq \theta_{l.2}$

The thresholds θ_i of each decision rule were optimized to achieve the best agreement (Dice-Score) with the segmentation of the *Necrosis* class as identified by BraTumIA. For this optimization, only the pre-operative examinations of the training set that contained necrotic voxels were included. The inclusion of only pre-operative data results from the fact that

resection cavities, namely the ones filled with blood, are often classified by BraTumIA as necrosis. There was a total of 7 MRSI examinations from 7 different patients meeting these requirements.

In the case of the *Solid Tumor* class, the aim was to detect very high *Cho/Cr* and/or *Cho/NAA* values that are associated with the parts of the tumor visible in sMRI. Consequently, the following rules were tested:

1. $Cho/Cr \geq \theta_{m.1}$ *AND* $Cho/NAA \geq \theta_{m.2}$
2. $Cho/Cr \geq \theta_{n.1}$ *OR* $Cho/NAA \geq \theta_{n.2}$
3. $Cho/Cr \geq \theta_{o.1}$
4. $Cho/NAA \geq \theta_{p.1}$

In this case, the thresholds were optimized to achieve the best agreement (Dice-Score) with the segmentation of the *MRI-visible tumor* (*Necrosis* U *E-Tumor* U *NE-Tumor*) as performed by BraTumIA. Similarly to the previous optimization, only pre-operative data from the training set was used. In this case, only examinations with *MRI-visible-tumor* voxels were included. There were 13 MRSI examinations from 13 different patients meeting these requirements.

The optimization of the thresholds was performed using Leave-One-Out-Cross-Validation (LOOCV), where every time one of the selected MRSI examinations were removed from the dataset used to optimize the thresholds and used to evaluate the performance.

10.3.7 BraTSIS validation and comparison with MRI-segmentation (BraTumIA)

The t-SNE method [Van Der Maaten 2008] was used to allow for the visualization of all the examples of the test set in a single 2D map. This map was then colored in various ways to highlight different aspects of the data:

1. Coloring/Labelling according to the closest cluster of the 5 different cluster solutions obtained with X-means only;
2. Coloring/Labelling according to the BraTSIS class of the closest cluster of the 5 different cluster solutions obtained with X-means;
3. Coloring/Labelling according to the BraTSIS class after classification of each individual spectra;
4. Coloring/Labelling according to the MRI-segmentation class.

Then, the median spectra and corresponding 5^{th}, 25^{th}, 75^{th} and 95^{th} percentile spectra were calculated for each class, using the examples of the test set. The same was performed for the values of each normalized feature used for clustering.

The MRS and MRI segmentations for all the cases of the test set, together with the maps of the different MRS features used in BraTSIS, were analyzed. Due to the high number of figures, the images of the second half of the test set were only included in the Supplementary Material. Finally, the relation between the MRI and MRS classes was further evaluated using the normalized confusion matrices containing the probabilities of each MRI class given a fixed MRI class, as well as the probabilities of each MRS class given a fixed MRI class.

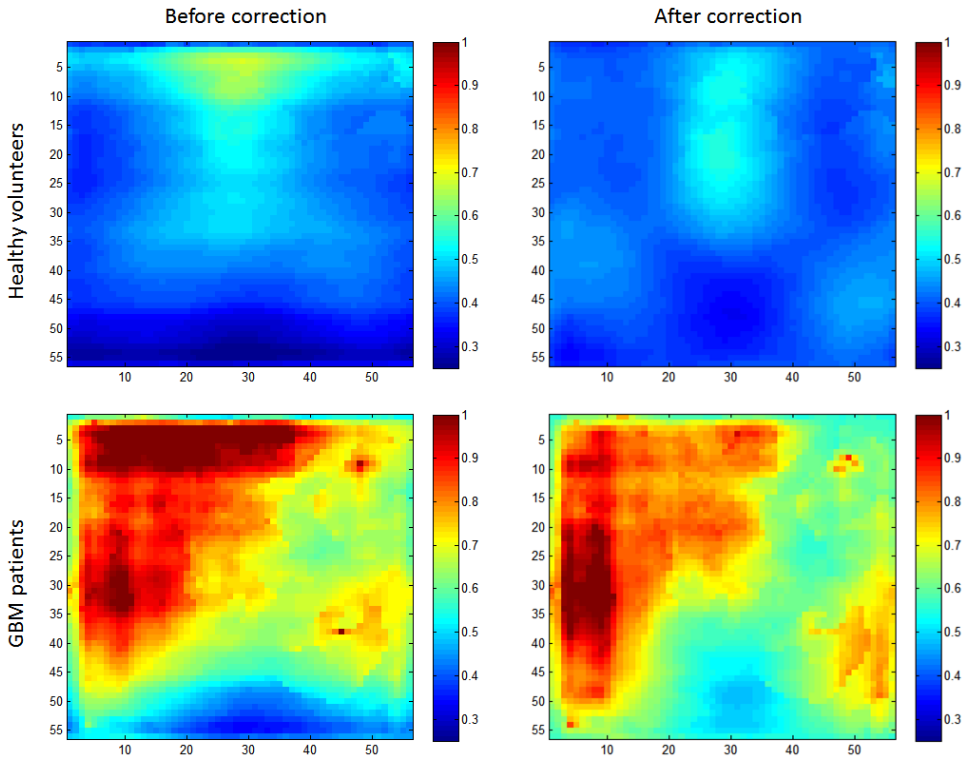

Figure 10.2: Average of localized Cho/NAA maps before and after selection profile correction, for healthy volunteers and GBM patients. The Cho/NAA maps were interpolated by a factor of 2, using zero-filling. Only data from within the PRESS-box is shown. The outer rows and columns of the PRESS box were removed before interpolation. The averaged maps are all transversally oriented, with the Left and Right sides of the patient inverted in the image (neuroradiological view).

10.4 Results

10.4.1 Selection-Profile Correction

Figure 10.2 shows the average *Cho/NAA* map for all healthy volunteers and GBM patients, before and after selection profile correction. The results show the high impact of the chemical shift artifact on the *Cho/NAA* maps. Without excitation profile and CSA-correction, normal brain tissue in the central frontal areas of the excited volume may be classified as tumor or infiltration. Similarly, if no correction is performed, tissue infiltration in the occipital areas may be classified as healthy brain tissue.

Besides these effects and despite the not perfect alignment of the different *Cho/NAA* maps, the average *Cho/NAA* map in GBM patients seems to indicate a higher prevalence of GBMs located in the right lobe than in the left lobe, confirming the results reported by Larjavaara et al.[Larjavaara 2007].

10.4.2 Clustering with X-means

The clustering performed with X-means produced an average of 52 clusters, in the range of 46 to 56 clusters. The clusters obtained in the first clustering solution, excluding the first 10 MRSI examinations of the training set, are shown in Figure 10.3. In that figure, the clusters are grouped by the corresponding BraTSIS class. The analysis of the clustering results was essential to determine the tissue-classes to use in BraTSIS, as well as to define the decision tree used to identify them.

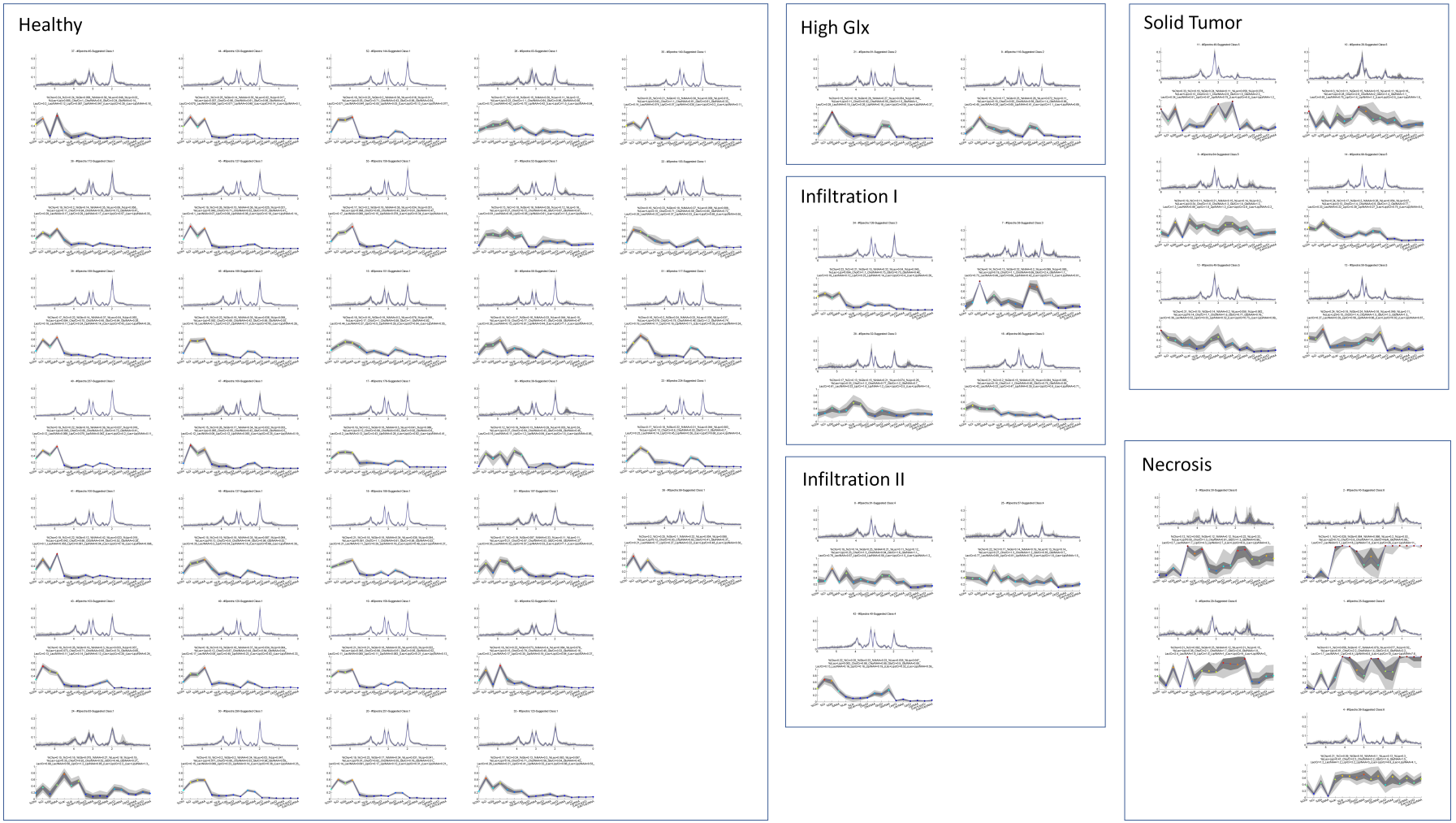

Figure 10.3: Magnitude spectra and normalized feature vectors for each cluster of the training set identified by X-means. The results shown correspond to the first X-means results, using the first four fifths of the training set. The different clusters are grouped by the corresponding BraTSIS class. For each cluster, the plots show the median values as well as 5th, 25th, 75th and 95th percentiles. For the normalized feature values, the coloring scheme used reflects the position of the value relative to the other values for that feature in the tested data set (dark red for values close to the maximum, and dark blue for values close the minimum value). In the titles of the plots, it is shown the cluster number, the number of spectra included in that cluster and the number of the BraTSIS class.

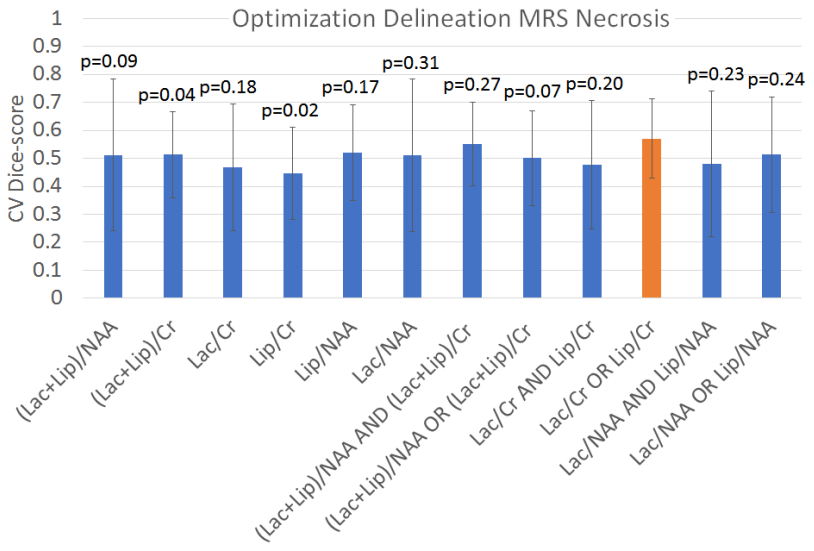

Figure 10.4: Dice scores between MRI Necrosis and MRS Necrosis defined using different features. In each case, the thresholds used were optimized to achieve the highest Dice-score with Necrosis as defined in MR imaging. The bar-plots show the standard deviation of the different results obtained with LOOCV. For each feature, it is shown the values of a paired t-test comparing the performance values obtained with that feature and the ones obtained with the one selected (Lac/Cr OR Lip/Cr). Considering the Bonferroni correction and a significance level of 0.05, there are no significant differences ($p<0.0042$) between the different features and the selected one.

10.4.3 Optimizing the detection of Necrosis and Solid Tumor

Figure 10.4 shows the agreement with the image segmentation of the decision rules using different features that were optimized to detect necrosis. The rule with the highest average Dice-score (0.571), and consequently selected to be used in BraTSIS, was "*Lac/Cr OR Lip/Cr*", with average $\theta_{j.1}$ and $\theta_{j.2}$ equal to 1.701 and 2.436. Despite this, several other tested decision rules lead to similar performances, as seen in the bar plot.

Similarly to Figure 10.4, Figure 10.5 shows the Dice-scores obtained with the different decision rules that were tested to detect *Solid Tumor*. As in the previous optimization, there were no significant performance differences between the different rules. Despite this, *Cho/Cr* alone provided the best average Dice-score (0.710). Moreover, the different optimal threshold values obtained with this parameter alone and in the "*Cho/Cr AND Cho/NAA*" and "*Cho/Cr OR Cho/NAA*" decision rules were considerably more stable than the optimal *Cho/NAA* values (coefficient of variation 8.4% vs 19.5%). These results suggest that *Cho/Cr* is more important to detect the solid part of the tumor than *Cho/NAA*. As a result, *Cho/Cr* alone was selected for the detection of *Solid Tumor*.

All the thresholds and performance measures related to the evaluated decision rules are included in the Supplementary Material in Table 10.2 and Table 10.3.

10.4.4 Infiltration I vs Infiltration II

The final decision rule to be fixed is the one related to the division of the *Infiltration* class into two subgroups with, assumingly, different malignancy levels. It is known that the two most commonly used metabolite ratios, *Cho/Cr* and *Cho/NAA*, are markers for tumor malignancy [Zeng 2011, Law 2003], but their relative importance may change at different stages of tumor development. The results of Cordova et al. [Cordova 2016] suggest that *Cho/NAA* is more

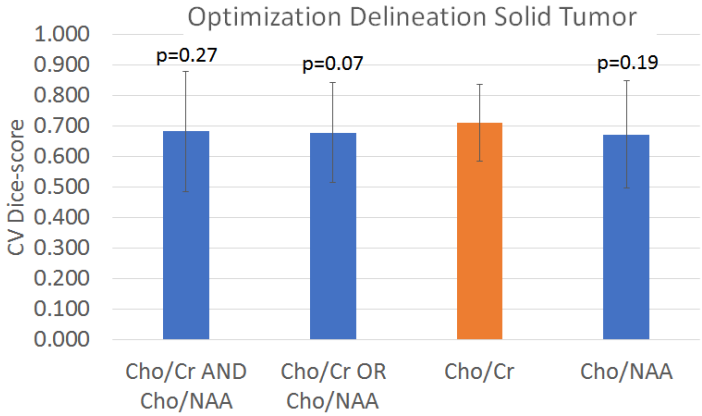

Figure 10.5: Dice scores between MRI-visible-tumor (Necrosis, E-Tumor, NE-Tumor), and MRS Solid Tumor defined using different features. In each case, the thresholds used were optimized to achieve the highest Dice-score with MRI-visible-tumor. The bar-plots show the standard deviation of the different results obtained with LOOCV. For each feature, it is shown the values of a paired t-test comparing the performance values obtained with that feature and the ones obtained with the selected one (Cho/Cr). Considering the Bonferroni correction and a significance level of 0.05, there are no significant differences ($p<0.0125$) between the different features and the selected one.

suitable to detect tumor infiltration than *Cho/Cr*. On the other hand, the results obtained with the optimization of the detection of *Solid Tumor* show that *Cho/Cr* is better at detecting the parts of the tumor visible in MRI. Considering this, it is expected that the division of the two subtypes of Infiltration occurs in a region where the importance of both features is similar. Consequently, the decision rule selected to separate these two subgroups considered both features:

$$Cho/Cr \geq \theta_{3.1} \; OR \; Cho/NAA \geq \theta_{3.2} \tag{10.4}$$

, where $\theta_{3.1}$ is the midpoint between the *Cho/Cr* threshold selected to identify *Solid Tumor* and the 99.9^{th} percentile of *Cho/Cr* in healthy volunteers, and $\theta_{3.2}$ is the midpoint between the best *Cho/NAA* threshold to detect *Solid Tumor* using *Cho/NAA* alone, and the 99.9^{th} percentile of *Cho/NAA* in healthy volunteers. The use of an OR in the decision rule guarantees that all *Infiltration* spectra with high values of *Cho/Cr* and/or *Cho/NAA* are identified, i.e., classified as *Infiltration II*.

10.4.5 BraTSIS and selected class membership tests

The BraTSIS decision tree with the different selected class membership tests is presented in Figure 10.6. All decision threshold values θ_k that were used in BraTSIS are shown in Table 10.1. Here, the uncertainty on the selected values is shown differently depending on the method used to determine the optimal threshold. For the thresholds optimized using image segmentation as reference, the uncertainty is reported as the standard deviation over the different results obtained using cross-validation. On the other hand, for the values determined as the 99.9^{th} percentile in healthy volunteers, the value is reported as the associated 95% confidence intervals, calculated as described in the methods section.

10.4.6 BraTSIS validation and comparison with MRI-segmentation (BraTumIA)

Figure 10.7 provides an overview of the different classifications of the test dataset. All the feature vectors of the test dataset were converted to points in a 2D-space, using t-SNE

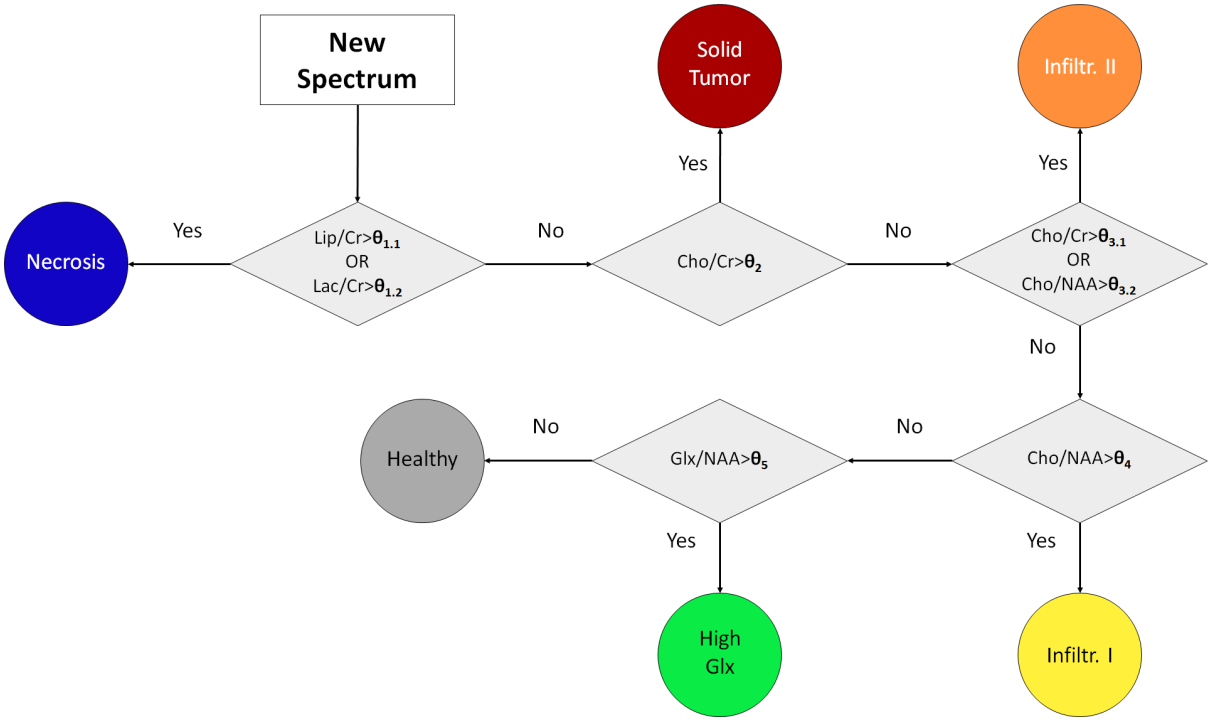

Figure 10.6: BraTSIS decision tree.

Table 10.1: All spectral features with their associated thresholds used in the current BraTSIS implementation. The first four values were obtained using LOOCV and therefore, the uncertainty is reported as the standard-deviation of the results obtained in the different iterations of the cross-validation. For the latter 3 features, the uncertainty is reported through the 95% CI associated with the 99.9^{th} percentile, calculated as described in the methods' section. The thresholds θ_2' and θ_4' were not directly used in BraTSIS but are essential to calculate the midpoint thresholds used to divide the two Infiltration type classes.

Threshold	Feature	Value	CV std	95%CI	Approach used
$\theta_{1.1}$	Lip/Cr	2.436	0.301	-	Optim. with Img. Segmentation
$\theta_{1.2}$	Lac/Cr	1.701	0.271	-	Optim. with Img. Segmentation
θ_2	Cho/Cr	1.342	0.021	-	Optim. with Img. Segmentation
θ_2'	Cho/NAA	1.105	0.053	-	Optim. with Img. Segmentation
$\theta_{3.1} = (\theta_2 + \theta_4')/2$	Cho/Cr	1.258	-	-	Midpoint
$\theta_{3.2} = (\theta_2' + \theta_4)/2$	Cho/NAA	0.898	-	-	Midpoint
θ_4	Cho/NAA	0.691	-	[0.684, 0.702]	99.9^{th} %ile healthy volunteers
θ_4'	Cho/Cr	1.174	-	[1.157, 1.207]	99.9^{th} %ile healthy volunteers
θ_5	Glx/NAA	0.749	-	[0.741, 0.760]	99.9^{th} %ile healthy volunteers

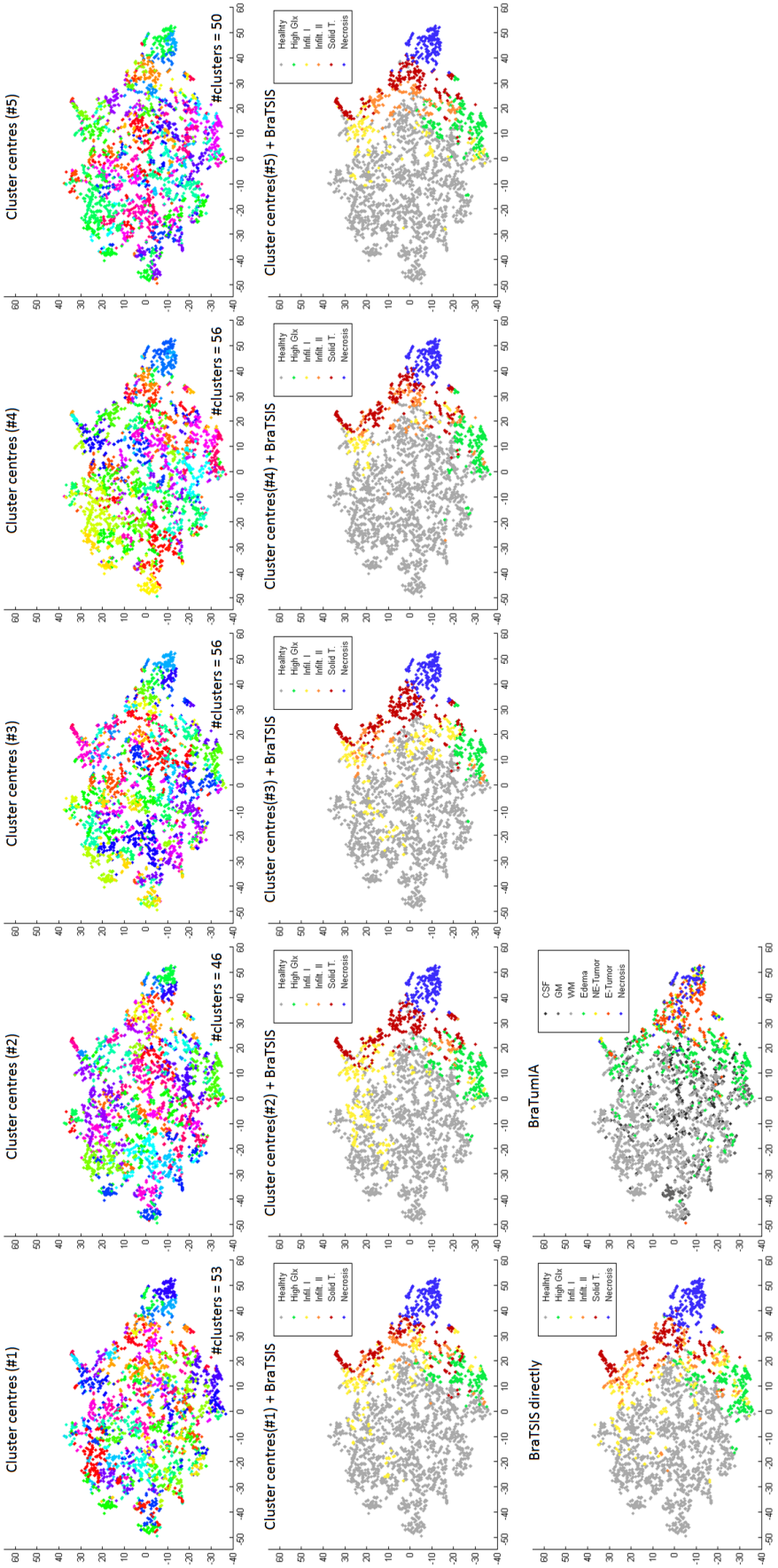

Figure 10.7: Maps showing all the spectra of the test set classified using: (**1.**) X-means results, (**2.**) X-means results + BraTSIS, (**3.**) BraTSIS alone and (**4.**) BraTumIA. The 2D-location of each point was calculated based on the values of the normalized features vectors, using t-SNE [Van Der Maaten 2008]. The maps of the first row show a different color depending on the closest cluster of each of the five X-means solutions (number of clusters of each solution shown in each map). Due to the elevated number of clusters, no legend was included for these maps. The second row shows the BraTSIS class corresponding to each of the cluster labels of the first row. Finally, the last two maps show the BraTSIS (applied directly) and BraTumIA classes of each example of the test set.

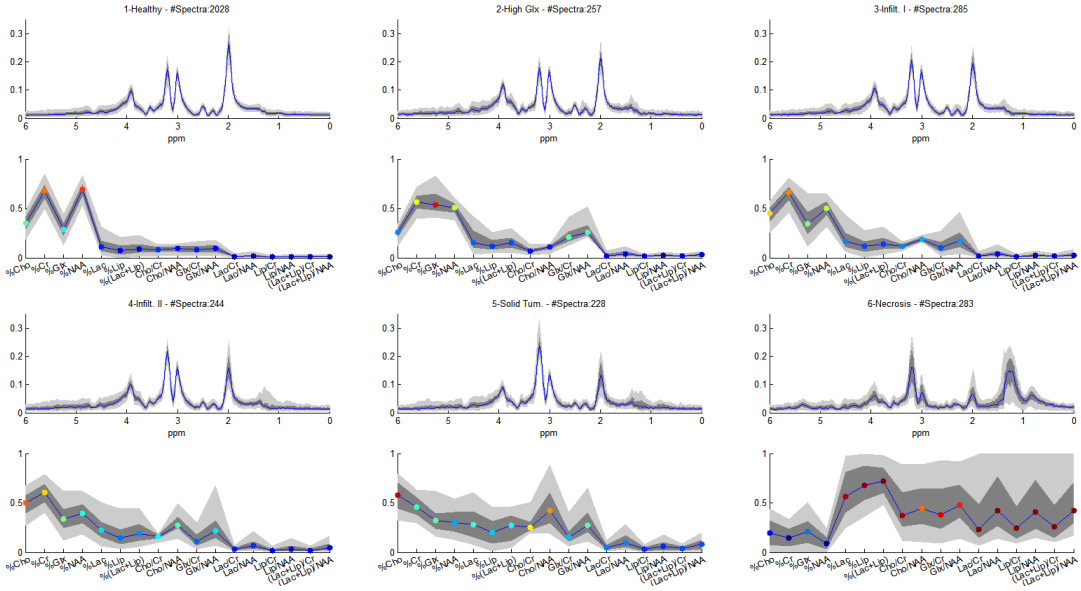

Figure 10.8: Magnitude spectra and normalized feature vectors for each class identified using BraTSIS directly on the test set. For each class, the plots show the median values as well as 5th, 25th, 75th and 95th percentiles. For the normalized feature values, the coloring scheme used reflects the position of the value relative to the other values for that feature in the tested data set (dark red for values close to the maximum, and dark blue for values close the minimum value). In the titles of the plots, the number of spectra that were assigned to each class is given.

[Van Der Maaten 2008]. The maps of the first row are colored based on the closest cluster number of each of the 5 cluster solutions. The second row shows the BraTSIS class of the clusters assigned to the first row. Finally, the last two maps show, on the left, the BraTSIS classification applying BraTSIS directly to the classification of each individual spectrum and, on the right, the BraTumIA classes of the corresponding regions in MRI.

The maps show that there are no well-isolated clusters in the test set, and besides not being shown here, the same is observed in the training set. This should explain the variability of the results of the first row of Figure 10.7. The results show that X-means succeeds in grouping similar spectra together, which is extremely valuable for the visualization of the different types of spectra existent in large datasets, as seen in Figure 10.3. Despite this, the variability of the clustering solutions and the elevated number of clusters indicates that X-means alone fails to identify reference groups that, once associated with a certain tissue-type, may directly be used to classify new spectra. This emphasizes the need for methods such as BraTSIS, based on clear rules that incorporate prior-knowledge in the classification of spectra. The maps show that BraTSIS succeeds in grouping similar spectra together while providing meaning to each spectrum/cluster.

Figure 10.8 shows the spectra and feature vectors from the test set grouped by BraTSIS class. In that figure, it is noticeable the low variance between the spectra of each group, a result that confirms the ability of BraTSIS to identify the major spectroscopic patterns of GBM patients.

The MRS and MRI segmentations of the first case of the test set are shown in Figure 10.9. In the same figure, different MR modalities (T_1, T_{1c}, T_2, FLAIR, SWI, and ADC) were also included. This example shows clearly that the Infiltration I class is found in regions with no alterations visible in MRI. Moreover, the combination of both segmentations may allow distinguishing infiltrated from non-infiltrated edema. In the same example, there is no agreement on the location of the necrosis.

Figure 10.10 shows the different segmentation results and metabolite maps of the first half

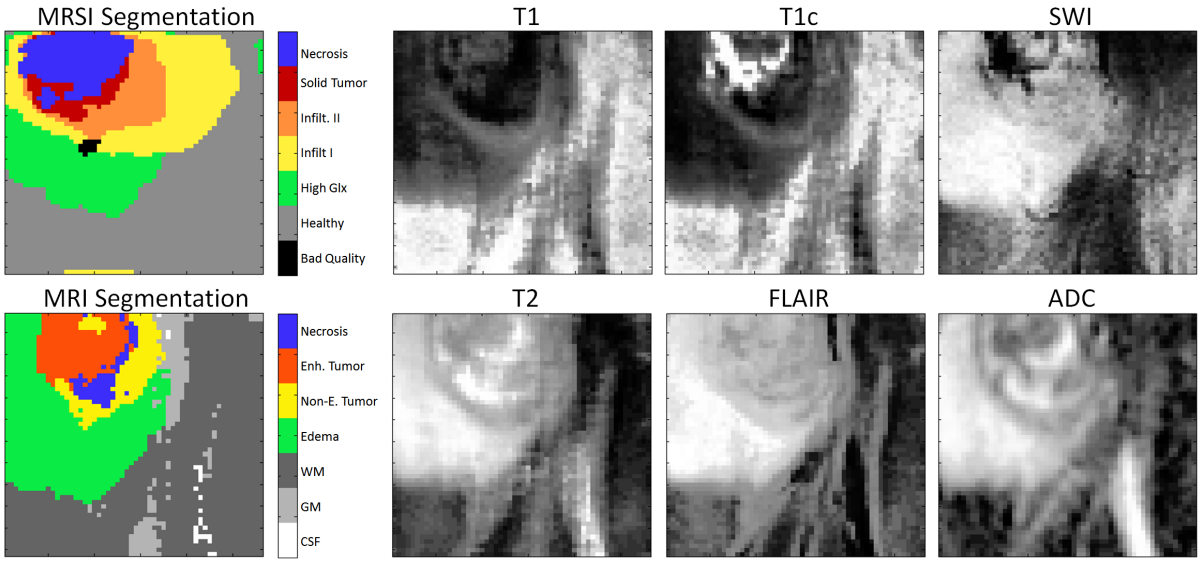

Figure 10.9: Example showing the segmentations obtained with BraTSIS and BraTumIA together with six different MR modalities with the same location and resolution as the segmentation maps. The pixel intensities of the 6 different MRI images show the mean value of the image voxels that correspond to each voxel of the interpolated MRSI grid. The example shown in the figure corresponds to the first example of the test set. The metabolite maps of this example are shown in the first row of Figure 10.10.

of the test set. The maps of the second half of the test set were included in the Supplementary Material. The different maps nicely illustrate the ability of BraTSIS to merge the information of the various metabolite ratios into a single map. The different examples show also interesting discrepancies between BraTumIA and BraTSIS, such as regions classified by BraTumIA as edema that show highly malignant features in MRS. The disparities between MRS and MRI segmentations reflect the differences in the information provided by these two techniques. The complementary aspect of MRI and MRS is particularly well analyzed in a recent paper [Lopez 2017] comparing different features extracted from T_{1c}- and T_2-weighted images and features extracted from MRSI. In that paper, the authors show, for instance, that Cho-derived maps show little correlation with MR imaging. Another study [Pirzkall 2004], where the use of MRSI to define target volumes for radiotherapy treatment planning was evaluated, found substantial differences in the spatial relationship between MRSI and MRI, further emphasizing the differences in the information provided by these two modalities.

The relations between the two classification methods are better understood with the normalized confusion matrix of Figure 10.11. Here we see that it is very unlikely for a spectrum to be classified as *Healthy* in spectroscopy and not healthy (*Edema*, *NE-Tum.*, *E. Tum.*, *Necrosis*) in MRI. On the other hand, about 30% of the spectra classified as *Infiltration I* and 21% of the spectra classified as *Infiltration II*, are found in *WM*.

Another interesting result is that the *High Glx* class is often found in *Edema*, which suggests an association between these two classes. The results show that 82.5% of the necrosis voxels identified in MRI are classified as necrosis in MRS, but only 30% of voxels classified as necrotic in MRS are identified as such by BraTumIA. This discrepancy may be explained by the under-segmentation of necrosis performed by BraTumIA previously described by Velasquez et al [Rios Velazquez 2015].

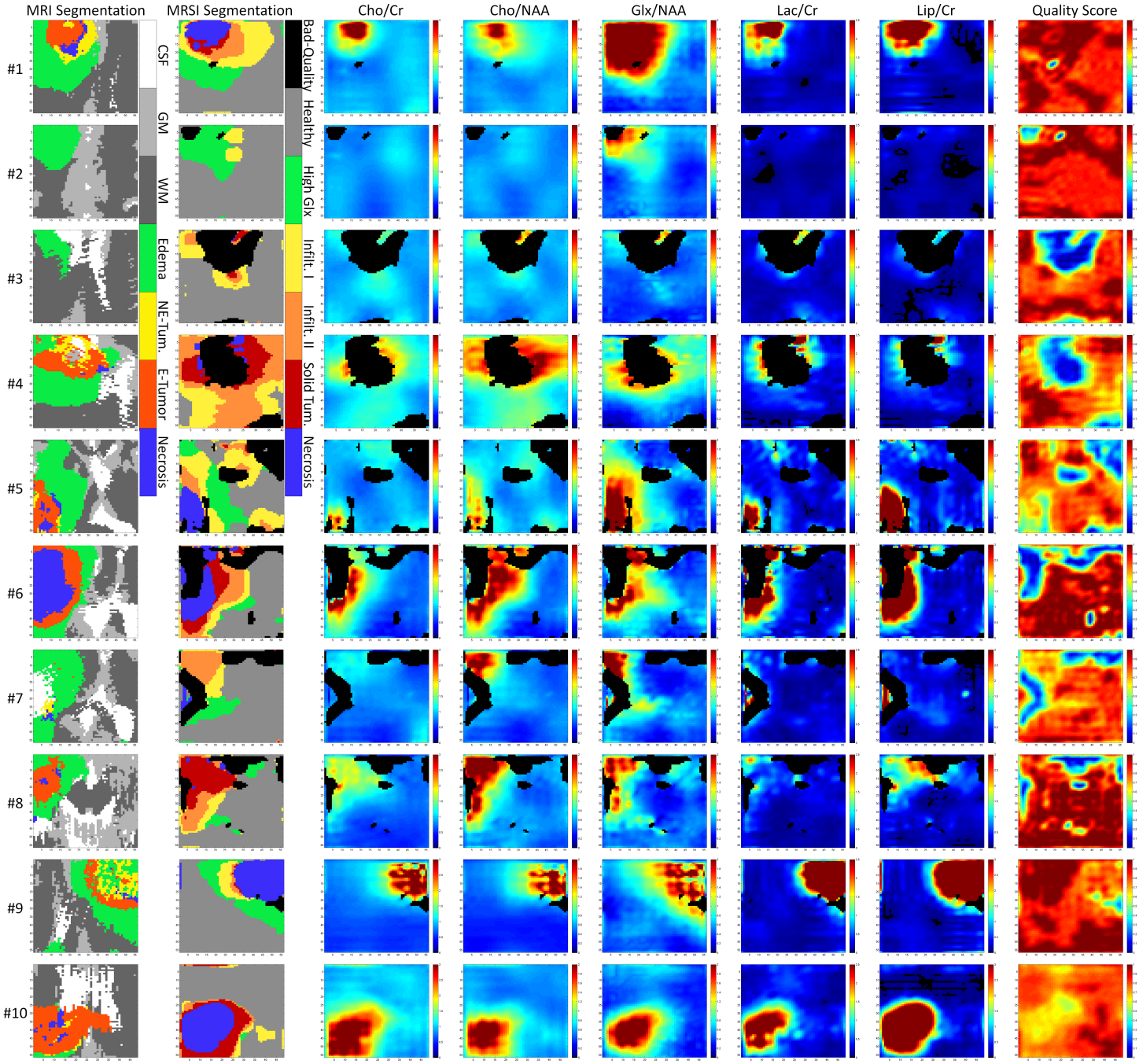

Figure 10.10: Segmentation results and metabolite maps for the first 10 MRSI examinations of the test set. The other 10 examinations were included in the Supplementary Material. The last column shows the quality score maps, representing the predicted probability that the quality of each spectrum is accepted by an expert rater [Pedrosa de Barros 2016b]. The regions of the MRS maps with quality scores below 0.5 were discarded and are shown in black. The color scales of both segmentations are shown next to the corresponding maps.

	CSF	GM	WM	Edema	NE-Tum.	E-Tum.	Necrosis
Healthy	0.129	0.169	0.662	0.036	0.001	0.003	0.000
High Glx	0.202	0.179	0.175	0.420	0.000	0.016	0.008
Infilt. I	0.161	0.196	0.298	0.326	0.011	0.007	0.000
Infilt. II	0.152	0.127	0.209	0.361	0.057	0.078	0.016
Solid Tum.	0.053	0.061	0.132	0.342	0.048	0.316	0.048
Necrosis	0.028	0.018	0.011	0.102	0.095	0.445	0.300

	Healthy	High Glx	Infilt. I	Infilt. II	Solid T.	Necrosis
CSF	0.628	0.125	0.110	0.089	0.029	0.019
GM	0.692	0.093	0.113	0.063	0.028	0.010
WM	0.863	0.029	0.055	0.033	0.019	0.002
Edema	0.154	0.231	0.199	0.188	0.167	0.062
NE-Tum.	0.035	0.000	0.053	0.246	0.193	0.474
E-Tum.	0.026	0.017	0.009	0.083	0.314	0.550
Necrosis	0.010	0.019	0.000	0.039	0.107	0.825

Figure 10.11: Normalized confusion matrices. The left matrix shows the probabilities of each MRI segmentation class for the voxels of a fixed MRS segmentation class. The right matrix shows the opposite: the probabilities of each MRS segmentation class for each MRI segmentation class. The results were calculated after applying BraTSIS directly to the interpolated test set.

10.5 Discussion

10.5.1 Automatic Quality Control and Selection Profile Correction are essential pre-processing steps

The results shown in the previous section confirm the feasibility of using a decision-tree based protocol to perform meaningful segmentations of brain tumor MRSI examinations in a simple and straightforward manner. Moreover, the development process of BraTSIS provided valuable insights into multiple metabolite ratios that are important for the characterization of brain tumors. For the successful development of BraTSIS, two processing steps were essential: Automatic Quality Control and Selection Profile Correction. Bad quality data, if not removed, prevents meaningful data exploration. The quality of the signals was assessed using SpectrIm's quality control [Pedrosa de Barros 2016b] algorithm, which can assess every grid almost instantaneously and with the same accuracy as a MRS-expert. Not only this saves a considerable amount of time compared to the manual assessment of the data, but also prevents subjective assessments of the quality of the signals, and enables fully automatic MRS data processing. With respect to Selection Profile Correction, which is a step that is often omitted in MRS data processing, the results demonstrated the negative impact that it can have on one of the most commonly used metabolite ratios: *Cho/NAA*. The step used for correction just requires a reference MRSI-scan of a phantom filled with appropriate metabolite solutions and is able to reduce significantly the impact of the CSA and of the non-perfect selection pulse profiles. More advanced corrections may provide further elimination of these effects. However, these artifacts are solved best at the pulse sequence level. MRSI pulse sequences suffering from these undesired effects are broadly used in clinical practice, emphasizing the need for such corrections. Consequently, there is an urgent need for such correction methods in the manufacturers' processing tools for the clinical routine use of MR spectroscopy.

10.5.2 Further research is needed to characterize the tissue properties associated with the High Glx class

The *High Glx* class originated from the clear identification by X-means of patterns showing overall healthy appearance but with significantly higher levels of *Glx* than other clusters. Despite its clear spectroscopic features, several questions are left open regarding the meaning of this class. Most of the *Edema* class seems to be divided in either *High Glx* or *Infiltration*, which may suggest that *High Glx* corresponds to non-infiltrated edema. Despite this, increased levels of Glx are associated with tumor growth [Takano 2001]. It is, therefore, necessary to understand whether the High Glx region corresponds mainly to peritumoral tissue under stress that has a minimal concentration of infiltrative tumor cells, or if this class captures the first

signs of tumor infiltration. Consequently, a future biopsy/MRS study evaluating the relation between *Glx/NAA* and tumor cellularity is necessary. The name "*High Glx*" chosen for this class reflects the need for a better understanding of the tissue characteristics associated with this metabolic pattern.

10.5.3 Comparison between hard and soft MRSI classification needs to be performed

BraTSIS is a hard classifier, i.e. assigns only one class to each spectrum. Soft classifiers, on the other hand, do not consider class membership as a binary decision and use continuous values (often a probability distribution) to describe the proximity between a given example and each class. The use of hard boundaries between classes may prevent interesting analysis related with elevated levels of *Glx* in other classes than *High Glx*. Similarly, the analysis of the overlap between Solid Tumor and Necrosis features could potentially allow a better description of the tumor core. Despite these possible drawbacks, the definition of well-defined classes with sharp boundaries reduces ambiguities and should benefit the standardization of MRSI analysis. The trade-offs between hard and soft classifications should be evaluated in a further study.

10.5.4 Cho/NAA and Cho/Cr are complementary features for tumor detection

Another interesting aspect that emerged from the development of BraTSIS was the comparison between Cho/NAA and Cho/Cr. Interestingly, Cho/Cr seemed to contain more information to identify the regions of the tumor visible in MRI (see results Figure 10.5). On the other, from previous research [Cordova 2016] it is known that Cho/NAA allows to better detect tumor infiltration. These results show that both features complement each other on the characterization of the different tumor regions of GBM.

10.5.5 Classification of spectroscopic patterns requires inclusion of prior-knowledge

Despite not being extensively discussed in this paper, the cluster analysis that laid the foundations for BraTSIS showed also that, in the analyzed dataset, there are no clear boundaries between the different types of spectra, a property that is well visible in the maps of Figure 10.7. This is most likely a result of the low resolution of the MRSI data. The low-resolution leads to spectra of mixed sources (partial-volume-effects) that fill the regions of the feature-space between clusters, connecting them. Alternatively, it is possible that the lack of clear cluster boundaries just reflects a continuous evolution of the metabolic patterns between healthy brain tissue and tumor tissue or necrosis, with no demarked stages. To test this hypothesis, it would be necessary to evaluate to apply BraTSIS using high-resolution MRSI data, using sequences such as EPSI [Mulkern 2001] or SPICE [Ma 2016, Lam 2015]. The lack of clear cluster boundaries causes the variability of the results of X-means seen in the first row of Figure 10.7. Apart from this, BraTSIS showed that it is possible to robustly identify different classes of spectra in case prior-knowledge is included in the classification. In this case, prior-knowledge corresponds not only to the class membership tests optimized using MRI segmentation or the abnormality thresholds derived from healthy volunteers, but also to everything that supported the decisions regarding the definition of the different classes, the structure of the decision-tree, and the choice of the features to be tested for each class membership test. The inclusion of prior-knowledge

is essential for the translation of the different spectroscopic patterns into clinically relevant information that can be used in the management of brain tumor patients.

10.5.6 BraTSIS may allow identification of tumor infiltration and improve delineation of tumor borders

The segmentation results of Figure 10.10 (and Figure 10.13 of the Supplementary Material) show that BraTSIS identifies regions that have features associated with tumor infiltration in MRS but the appearance of normal brain tissue or edema in conventional imaging. This mismatch between BraTSIS and BraTumIA is seen in almost every case of the test set. *Cho/NAA* alone has shown to be able to identify regions with high tumor cellularity not visible with sMRI [Cordova 2016, Guo 2012, Stadlbauer 2004], which supports the belief that these regions identified by BraTSIS correspond to tissue with elevated tumor cellularity and that BraTSIS may be used to improve the delineation of brain tumor borders. Despite the strong evidence from the previous literature that supports this idea, further validation with histopathology is required. The inclusion of *Cho/NAA* maps to delineate GBM lesions and plan Gamma Knife stereotactic surgery has shown to improve patient survival in a Phase II trial conducted at the Case Western Reserve University Kettering, Ohio, in 2012 [Einstein 2012].

10.5.7 BraTSIS improves detection of tumor necrosis

In GBM patients, the existence of a necrotic core is associated with poor prognosis [Burger 1987, Barker 1996]. *Necrosis* can be easily identified in MRS by the high levels of lipids and/or lactate and low levels of other metabolites [Vigneron 2001, Horská 2010]. The examples number #5, #9 and #10 of Figure 10.10, show 3 tumors where BraTumIA identifies only a few voxels as necrosis but BraTSIS shows relatively big necrotic cores, with high levels of *Lip/Cr* and *Lac/Cr*. Moreover, the latest MICCAI BraTS'17 challenge (`www.med.upenn.edu/sbia/brats2017.html`) (Multimodal Brain Tumor Segmentation Challenge) does no longer distinguish between Necrosis and NE Tumor, which reflects the known difficulties in detecting tumor necrosis using sMRI described by Velasquez and Meier et al. [Rios Velazquez 2015]. This further emphasizes the importance of MRSI brain tumor segmentation for the complete characterization of brain tumors.

10.5.8 BraTSIS may help to identify the most malignant parts of the tumor

Several studies [Zeng 2011, Law 2003] have shown the ability of MRS to distinguish different tumor grades, based on *Cho/Cr* and *Cho/NAA*. Furthermore, MRS has been shown to help to identify foci of higher tumor malignancy in stereotactic brain biopsy [Hermann 2008, Son 2001], which may avoid wrong tumor grading due to sampling errors. BraTSIS considers different classes whose features are associated with different malignancy levels, namely Solid Tumor, Infiltration II and Infiltration I. Therefore, BraTSIS may help to identify the most malignant parts of the tumor and provide relevant information for guiding stereotactic brain biopsy.

10.6 Conclusion

The BraTSIS protocol presented here combines several feature threshold rules in a simple decision-tree. Despite its simplicity, BraTSIS effectively identifies the main types of spectra seen in MRSI examinations of GBM patients. The results showed the potential of this approach to

Table 10.2: Feature thresholds and performance measures obtained after optimizing different MRS features to obtain the highest Dice scores with Necrosis as segmented by BraTumIA. (std – standard deviation, Acc – accuracy, Sens – Sensitivity, Spec – Specificity, Prec – Precision, F1 – F1-score)

	$\theta_{X.1}$	std	$\theta_{X.2}$	std	Acc	std	Sens	std	Spec	std	Prec	std	F1	std
(Lac+Lip)/NAA	2.521	0.271			0.921	0.271	0.715	0.271	0.939	0.271	0.481	0.271	0.512	0.271
(Lac+Lip)/Cr	3.158	0.440			0.911	0.050	0.764	0.284	0.927	0.059	0.489	0.249	0.513	0.153
Lac/Cr	1.502	0.264			0.926	0.040	0.644	0.351	0.942	0.051	0.430	0.268	0.468	0.227
Lip/Cr	2.032	0.300			0.903	0.046	0.741	0.310	0.923	0.054	0.435	0.264	0.446	0.166
Lip/NAA	1.523	0.232			0.915	0.050	0.812	0.275	0.934	0.054	0.492	0.260	0.521	0.171
Lac/NAA	1.127	0.071			0.933	0.043	0.635	0.346	0.948	0.050	0.520	0.352	0.511	0.272
(Lac+Lip)/NAA AND (Lac+Lip)/Cr	3.055	0.416	2.300	0.098	0.928	0.042	0.723	0.258	0.948	0.044	0.540	0.233	0.550	0.149
(Lac+Lip)/NAA OR (Lac+Lip)/Cr	7.016	6.628	2.734	0.924	0.919	0.040	0.723	0.258	0.938	0.047	0.479	0.264	0.502	0.169
Lac/Cr AND Lip/Cr	1.354	0.140	0.251	0.565	0.929	0.040	0.644	0.351	0.945	0.052	0.448	0.288	0.477	0.229
Lac/Cr OR Lip/Cr	1.701	0.271	2.436	0.300	0.926	0.046	0.759	0.233	0.939	0.053	0.523	0.213	0.571	0.142
Lac/NAA AND Lip/NAA	1.127	0.071	0.151	0.370	0.928	0.045	0.601	0.363	0.950	0.051	0.529	0.364	0.480	0.261
Lac/NAA OR Lip/NAA	1.135	0.065	1.773	0.467	0.926	0.046	0.701	0.281	0.936	0.056	0.543	0.309	0.513	0.207

Table 10.3: Feature thresholds and performance measures obtained after optimizing different MRS features to obtain the highest Dice scores with the MRI-visible-tumor as segmented by BraTumIA. (std – standard deviation, Acc – accuracy, Sens – Sensitivity, Spec – Specificity, Prec – Precision, F1 – F1-score)

	$\theta_{X.1}$	std	$\theta_{X.2}$	std	Acc	std	Sens	std	Spec	std	Prec	std	F1	std
Cho/Cr AND Cho/NAA	1.308	0.060	0.747	0.159	0.891	0.082	0.771	0.247	0.903	0.111	0.686	0.265	0.683	0.198
Cho/Cr OR Cho/NAA	1.575	0.199	1.212	0.134	0.883	0.083	0.784	0.213	0.894	0.117	0.652	0.237	0.679	0.164
Cho/Cr	1.342	0.021			0.891	0.074	0.838	0.163	0.893	0.109	0.673	0.222	0.710	0.125
Cho/NAA	1.105	0.053			0.892	0.076	0.730	0.245	0.915	0.102	0.699	0.252	0.673	0.177

(a.) identify tumor infiltration and improve the tumor delineation performed using sMRI, (b.) identify tumors with necrotic cores, and (c.) help to identify the most malignant parts of brain tumors. The simplicity of the approach facilitates its use as a standard for MRSI interpretation and allows it to be easily extended and improved in the future to maximize the utility of MRSI for brain tumor characterization.

10.7 Acknowledgements

This work was funded by the EU Marie Curie FP7-PEOPLE-2012-ITN project TRANSACT (PITN-GA-2012-316679) and the Swiss National Science Foundation (project number 140958).

10.8 Supplementary Material

10.8.1 A: Detailed results of the optimization of the detection of Necrosis and Solid Tumor

Table 10.2 and Table 10.3 show the results of the optimization of different decision rules for the detection of *Necrosis* and *Solid Tumor*.

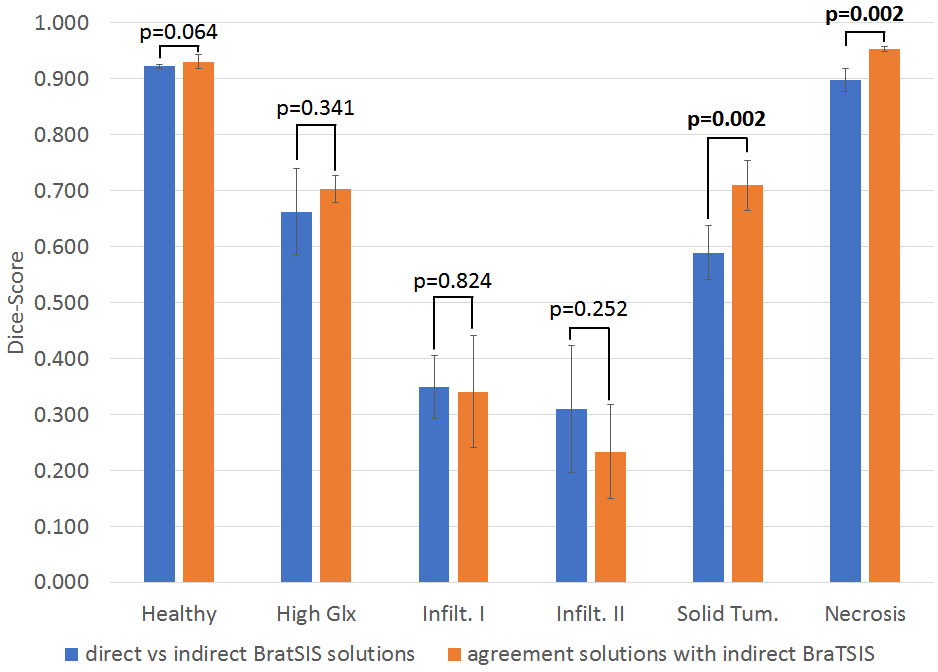

Figure 10.12: Comparison between the classification of spectra of the test set directly using BraTSIS to classify each individual spectrum, and indirectly, using BraTSIS to classify the clusters identified using X-means. The indirect approach was tested using the 5 different cluster solutions. The statistical significance of the differences between the agreement of indirect approaches with the direct approach and the agreement between the different indirect solutions was evaluated using a two-tailed t-test considering 2 samples with unequal variance.

10.8.2 B: Classification of individual spectra using BraTSIS

Figure 10.12 shows the comparison between the application of BraTSIS to classify individual spectra (direct BraTSIS) and classification by copying the BraTSIS class of the closest cluster (indirect BraTSIS). The test was performed applying both approaches to classify the spectra of the test set. The results show that there are only small significant differences between these approaches in the detection of *Solid Tumor* and *Necrosis*. Interestingly, the extreme classes, *Healthy* and *Necrosis*, are the ones with the highest overall agreement. The relatively low Dice scores seen in the other classes demonstrate the variability of the results caused by the differences in the cluster solutions used in the indirect approach. This further supports the interest in using BraTSIS directly to classify each spectrum without relying on clustering solutions for intermediate cluster assignment.

10.8.3 C: Segmentation results of the second half of the test set

Figure 10.13 shows the segmentation results of the second half of the test set.

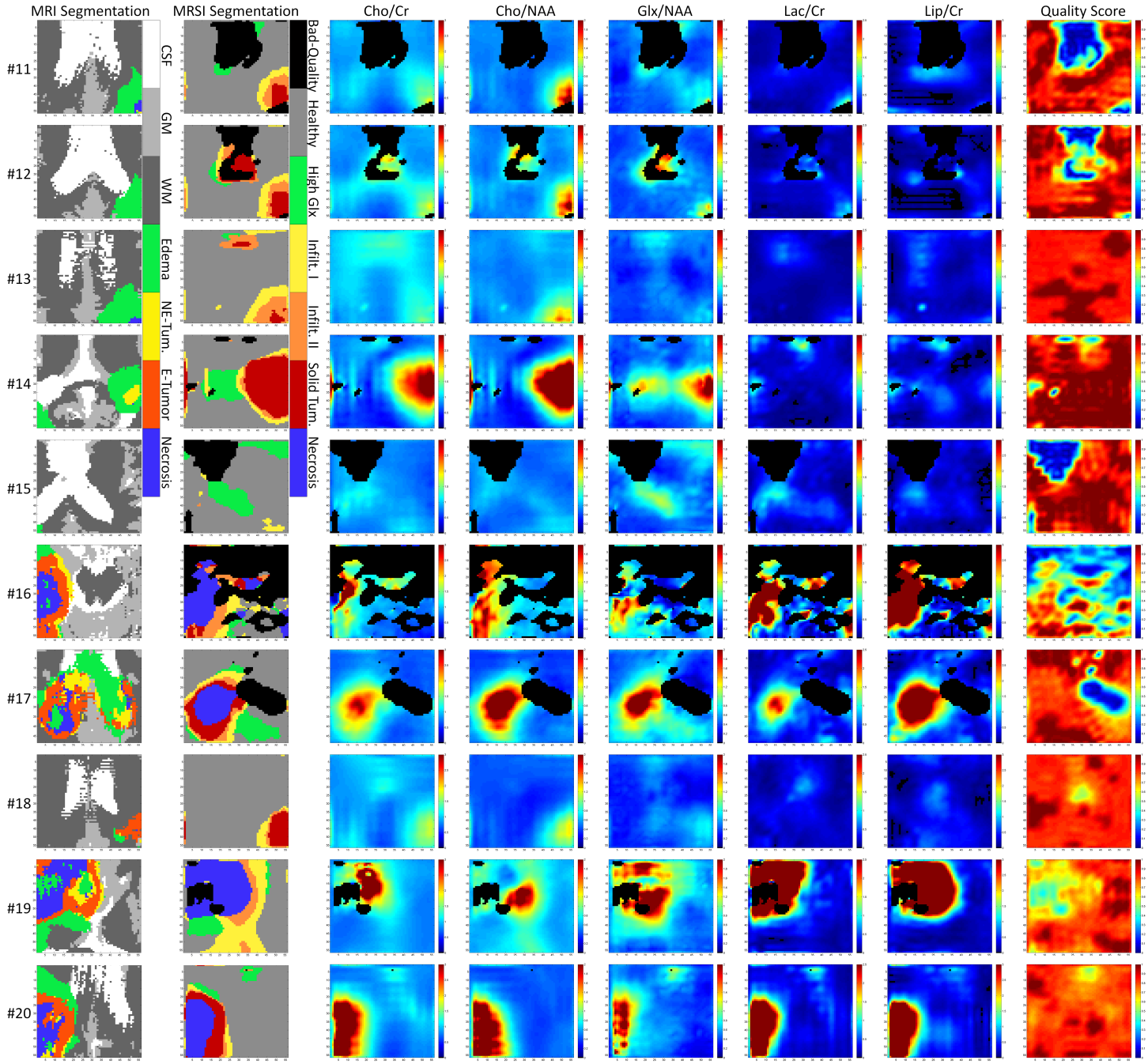

Figure 10.13: Segmentation results and metabolite maps for the second half of the test set. The last column shows the quality score maps, representing the predicted probability that the quality of each spectrum is accepted by an expert rater [Pedrosa de Barros 2016b]. The regions of the MRS maps with quality scores below 0.5 were discarded and are shown in black. The color scales of both segmentations are shown next to the corresponding maps.

11 | On the relation between MR Spectroscopy features and the distance to the MRI-visible Solid Tumor in GBM patients

Authors: Nuno Pedrosa de Barros, Raphael Meier, Martin Pletscher, Samuel Stettler, Urspeter Knecht, Evelyn Herrmann, Philippe Schucht, Mauricio Reyes, Jan Gralla, Roland Wiest, Johannes Slotboom.

The content of this chapter was *submitted* for publication to the journal *Magnetic Resonance in Medicine*.

11.1 Abstract

Purpose: To improve the detection of peritumoral changes in GBM patients by exploring the relation between MRSI information and the distance to the solid tumor volume (STV) defined using structural MRI (sMRI).
Methods: 23 MRSI studies (PRESS, TE 135ms) acquired from different patients with untreated GBM were used in this study. For each MRSI examination, the STV was identified by segmenting the corresponding sMRI images using BraTumIA, an automatic segmentation method. The relation between different metabolite ratios and the distance to STV was analyzed. A regression forest was trained to predict the distance from each voxel to STV based on 14 metabolite ratios. Then, the trained model was used to determine the Expected Distance to Tumor (EDT) for each voxel of the MRSI test data. EDT maps were compared against sMRI segmentation.
Results: The features showing abnormal values at the longest distances to the tumor were: %NAA, Glx/NAA, Cho/NAA and Cho/Cr. These 4 features were also the most important for the prediction of the distances to STV. Each EDT value was associated with a specific metabolic pattern, ranging from normal brain tissue to actively proliferating tumor and necrosis. Low EDT values were highly associated with malignant features such as elevated Cho/NAA and Cho/Cr.
Conclusion: The proposed method enables the automatic detection of metabolic patterns associated with different distances to the STV border and may assist tumor delineation of infiltrative brain tumors such as GBM.

11.2 Introduction

Glioblastoma Multiforme (GBM) is the most aggressive type of primary brain tumor and is characterized by an extensive infiltration of tumor cells into the tissues surrounding the tumor [Claes 2007]. Structural MRI (sMRI) is the imaging technique of reference for the management

of GBM patients [Weller 2014, Niyazi 2016] but fails to show the full extent of the infiltrative tumor part [Yamahara 2010], complicating the treatment of this disease.

In order to improve the delineation of these tumors, several research groups [Cordova 2016, Guo 2012, Stadlbauer 2004, Ganslandt 2005] have used Magnetic Resonance Spectroscopy (MRS), a non-invasive technique that provides metabolic in-vivo information of the tissues examined and complements sMRI in the characterization of brain tumors [Hollingworth 2006]. Results have shown that MRS is able to detect regions with tumor cells outside the tumor limits defined in T_2-weighted images [Ganslandt 2005]. Improvement in the delineation of glioblastoma has the potential to improve patient treatment and increase the overall survival of these patients, as indicated by a phase II clinical trial [Einstein 2012] where a significant increase in GBM patient's overall survival was observed after incorporating MRS in the planning of radiosurgery. Similarly, during brain surgery, the increased delineation accuracy that can be achieved by using 5-aminolevulinic acid (5-ALA) tumor fluorescence [Schucht 2014], translates into a more effective tumor resection with increased survival to the patients [Aldave 2013].

Despite these promising results on the detection of tumor infiltration using MRSI, in the above-mentioned work, as in current clinical practice, only a very limited amount of the information provided by MRS is used. More precisely, the improvement in tumor detection is usually achieved simply by thresholding the metabolite ratio Cho/NAA. Therefore, it is important to evaluate alternative approaches that, using more information, may further improve tumor delineation and detection of tumor infiltration.

Several researchers have developed methods that, instead of analyzing each metabolite or metabolite ratio individually, aim at identifying metabolic patterns associated with different tissues [Preul 1998, Sajda 2004, Du 2004, Du 2008, Su 2008, Luts 2009, Ortega-Martorell 2012a, Lu 2013, Li 2013b, Li 2013a, Yang 2015, Raschke 2015, Mocioiu 2016, Pedrosa de Barros 2015]. These methods can be divided into two main groups: classification methods [Preul 1998, Luts 2009, Lu 2013, Yang 2015] and (blind) source separation methods [Sajda 2004, Du 2004, Du 2008, Su 2008, Ortega-Martorell 2012a, Li 2013b, Li 2013a, Raschke 2015, Mocioiu 2016, Pedrosa de Barros 2015]. In the first group the analysis of each spectrum corresponds to the classification of each spectrum into one of a set of predefined classes. Therefore, the tumor boundary is simply the region delimited by the voxels classified as tumoral tissue. In contrast, the second group presents a set of methods that model each spectrum as a linear combination of different reference spectra of predefined tissue-types (e.g. necrosis, tumor and normal brain tissue). By doing so, tumor delineation can be performed based on the estimations of the tumor-tissue contribution in each spectrum. This type of methods is more suitable for delineating GBM since they are prepared to deal with diffuse tumor boundaries and with partial-volume effects that originate from the low spatial resolution of MRSI data. In both cases, the results can be made easily interpretable by converting the output into nosologic images, i.e. images where each color represents a given tissue type [De Edelenyi 2000].

The two types of methods presented above enable a better use of the MRSI data for tumor delineation, however, their availability in clinical practice is still very limited. One of the main challenges for the incorporation of such methods into clinical routine is that their results depend on the set of classes or reference spectra used, which not only may vary considerably depending on the dataset and methodology used, but also usually require a non-trivial decision regarding the number of classes or reference spectra that should be considered.

In this work, a novel approach to explore MRSI information for brain tumor delineation is presented. The method was developed considering the solid tumor volume (STV) identified in

sMRI as the core of the tumor. Moreover, it was assumed that the probability of finding tumor cells decreases as the distance to the STV increases. Under these assumptions, we explored the relation between metabolic patterns and the distance to the STV defined in sMRI. In this way, it may be possible to detect tumor infiltration by simply identifying metabolic signatures that are associated with close distances to the STV.

The work here presented starts with an analysis of the relation between individual metabolite ratios and the distance to Solid Tumor Volume (STV). For comparison, we looked also at the values of these metabolite ratios for the different MRI-segmentation classes (white-matter (WM), gray-matter (GM), cerebral-spinal-fluid (CSF), enhancing-tumor (E-Tumor), non-enhancing tumor (NE-Tumor) or necrosis). The analysis of the relation between each feature and the distance to STV was supported by an *MRS-Feature Spatial Distribution* (MRS-FSD) model that is introduced in this article.

After the analysis of the different features separately, a regression forest [Breiman 2001, Liaw 2002] was trained to predict the distance to STV based on the value of 14 metabolite ratios in each spectroscopic voxel. The relation between metabolic patterns and the distances predicted by the regression forest, or Expected Distances to Tumor (EDT), was analyzed. Finally, the contours defined by the EDT maps produced by the trained model were compared with sMRI segmentation.

11.3 Methods

11.3.1 Data

A total of 23 pre-operative MRSI examinations (PRESS, CHESS water suppression, TE=135 ms, TR=1500 ms) were collected from different GBM patients using two 1.5T Siemens scanners (models 'Aera' and 'Avanto', Siemens, Erlangen, Germany). The MRSI data had an original resolution of 12x12, with a voxel size of 13.33x13.33x15 mm, and was interpolated before DICOM storage to a resolution of 32x32 and a voxel size of 5x5x15 mm. All spectra from outside the PRESS-box, as well as the border voxels of the excited PRESS volumes, were discarded as these voxels are not properly excited. Moreover, the border voxels of the PRESS volume are usually highly affected by the chemical shift displacement artifact and cannot be corrected using methods such as the selection profile correction discussed in the next section. After removing the outer voxels, the dataset contained a total of 4452 spectra. Besides spectroscopy, the protocol included acquisition of T_1-weighted (MP RAGE, TE=2.67ms; TR=1580ms, TI=900ms; 1x1x1 mm), T_{1c}-weighted post-contrast (MP RAGE, TE=4.57ms; TR=2070ms, TI=1100ms; 1x1x1 mm), T_2-weighted (SPACE, TE=380ms, TR=3000ms), and FLAIR MRI (TE=88ms, TR=8000ms, TI=2500ms).

Apart from the data collected from GBM patients, 7 MRSI recordings were acquired from 3 healthy volunteers using the same acquisition parameters. This healthy control dataset contained a total of 1372 spectra after exclusion of the outer-voxel row of the excited volume.

The measurements were performed in conformance to local and national ethical regulations, meaning that all patients gave their written informed consent that their data could be used for scientific purposes.

11.3.2 MRS Data Processing Pipeline

The MRSI data was processed using jMRUI's SpectrIm plugin (`www.jmrui.eu`) and included the following processing steps:

1. Residual water peak removal using HLSVD [den Boogaart 1994];
2. Automatic Quality Control [Pedrosa de Barros 2016b, Pedrosa de Barros 2016a, Pedrosa de Barros 2017a];
3. Frequency-Shift Correction;
4. Auto-phasing;
5. Quantification using QUEST [Ratiney 2004];
6. Selection profile correction [Pedrosa de Barros 2017b].

Quality Control was performed using a random forest model that was trained to reproduce the judgment of an expert. The criteria for rejection of the training data was the presence of visible signal artifacts and low SNR. The trained model outputs a probability of acceptance by an expert. Therefore, after assessing all spectra with the automatic classifier, all spectra with a probability of acceptance by an expert below 0.5 were removed from the dataset. After quality filtering, the dataset contained 3733 spectra, meaning that around 16% of the spectra (719 spectra) were classified as having unacceptable quality. For more details on automatic quality filtering the reader is referred to [Pedrosa de Barros 2016b, Pedrosa de Barros 2016a, Pedrosa de Barros 2017a, Pedrosa de Barros 2017b].

Quantification was performed using QUEST and a metabolite model that included the following metabolites: Cho, Cr, NAA, Glu, Gln, Lac, Lip1.3, Lip0.9.

Chemical-Shift Displacement artifact (CSDA) and RF selection pulse profile effects of the PRESS sequence were corrected using metabolite maps recorded of phantoms using the same acquisition parameters as the in-vivo measurements. This correction is essential for the analysis of MRSI, reducing the spatial variations of the relative metabolite content that originate from the non-ideal selection pulses. In-depth information on the selection profile correction is found in [Pedrosa de Barros 2017b].

11.3.3 Image Segmentation using BraTumIA

The sMRI images acquired for each MRSI examination were segmented into the following compartments: white-matter (WM), gray-matter (GM), cerebral-spinal-fluid (CSF), edema, non-enhancing tumor (NE-Tumor), enhancing tumor (E-Tumor) and necrosis. Solid Tumor Volume (STV) was defined as the union of NE-Tumor, E-Tumor, and necrosis.

The segmentation was performed fully automatically using BraTumIA [Porz 2014, Meier 2016]. BraTumIA is a machine learning-based segmentation method that was trained on an independent dataset of 54 pre- and postoperative MRI examinations of glioblastoma that were manually segmented by experienced neuroradiologists. For the automatic segmentation, the following modalities are required as input: T_1, T_{1c}, T_2, and FLAIR. A more detailed description of the algorithm and training data can be found in the study of Meier et al. [Meier 2016].

11.3.4 MRS features

Two types of features were analyzed: (I.) *%metabolite*, calculated as the ratio between the area of a given metabolite and the total sum of the areas of all metabolites, and (II.) conventional ratios between the areas of two metabolites. The areas were calculated by integration of the complete metabolite signal of the basis set after fitting with QUEST and were corrected for the selection profile. Considering this, the following 14 MRS features were analyzed in this paper: %NAA, %Cho, %Cr, %Glx, %Lac, %Lip, Cho/Cr, Cho/NAA, Glx/NAA, Glx/Cr, Lac/Cr, Lip/Cr, Lac/NAA and Lip/NAA. In these features, Lip stands for the sum of Lip1.3 and Lip0.9, and Glx for the sum of Glu and Gln.

11.3.5 Spectra and MRS features for different distances to the STV and different Image Segmentation Classes

For each spectroscopic voxel, the assigned MRI-segmentation class was selected based on majority voting, considering the voxels of the image segmentation mask that are contained within the spectroscopic voxel. Then, the distance from each voxel to the *STV* was calculated as the Euclidean distance to the nearest voxel assigned to one of the classes of the *STV (NE-Tumor*, *E-Tumor*, *Necrosis)*. The distance was calculated in the plane of the MRSI grid and rounded to the closest distance corresponding to an integer number of voxels.

The spectra of all spectroscopic voxels were grouped based on the rounded distance to STV and based on the segmentation class. All spectra were normalized using L_1-normalization, and, for each group (i.e. range of distances to STV or segmentation class), the median spectra and the 5^{th}, 25^{th}, 75^{th} and 95^{th} percentile spectra were calculated.

Similarly to the spectra, the 14 previously described features were also grouped based on the segmentation class and the distance to STV. For each MRSI examination, the different feature maps were spatially interpolated by a factor of 4 (zero-filling) to enable a finer analysis of the relation between each feature and the distance to *STV*. After interpolation, the final in-plane resolution was 1.25x1.25 mm. For each distance to *STV* and for each image segmentation class, the median values as well as the 5^{th}, 25^{th}, 75^{th} and 95^{th} percentiles were calculated. The same percentiles were calculated for all voxels of the healthy volunteers' dataset and used as a reference.

11.3.6 MRS-Feature Spatial Distribution (MRS-FSD) model

In order to compare the spatial properties of the different metabolite ratios, namely their relation with the distance to *STV*, the following *MRS-Feature Spatial Distribution (MRS-FSD)* model was introduced:

$$x_i(d) = K_i + A_i e^{-\frac{d}{\lambda_i}} \tag{11.1}$$

where d is the distance, and K_i, A_i and λ_i are the model parameters for the feature i. In this expression, K_i represents the value of the feature i when the distance to the tumor approaches infinity ($K_i = x_i(d = \infty)$). The fit parameter A_i (amplitude) represents the difference between the values found in the *STV* and the values found in regions far from the *STV* ($A_i = x_i(d = 0) - x_i(d = \infty)$). The distance parameter λ_i, or *spectral feature normalization length*, controls the characteristic rate at which the parameter normalizes as the distance to the tumor increases.

To facilitate the comparison between different features, the values of each feature were rescaled before fitting the model using the following expression:

$$x_i' = \frac{x_i - {}^{H}X_i^{5th}}{{}^{H}X_i^{95th} - {}^{H}X_i^{5th}} \tag{11.2}$$

where ${}^{H}X_i^{5th}$ and ${}^{H}X_i^{95th}$ correspond respectively to the 5th and the 95th percentiles of the same feature in healthy subjects, and x_i and x'_i correspond to the value of the feature in GBM patients before and after rescaling. The rescaling enables direct comparison of the different model parameters between different MRS features and allows for an easier interpretation of the results.

Following the above described feature rescaling, we define the abnormality distance of a given feature i, $D_{abn,i}$, as the distance at which each feature crosses the abnormality threshold. The abnormality threshold was defined as ${}^{H}X_i^{95th}$ for features that are increased in tumoral tissue relatively to normal brain tissue, and as ${}^{H}X_i^{5th}$ for features that are decreased in tumoral tissue. Therefore, $D_{abn,i}$ is calculated as:

$$D_{abn,i} = \begin{cases} -\lambda_i \cdot \log(\frac{1-K_i}{A_i}) & , A_i > 0 \\ -\lambda_i \cdot \log(\frac{-K_i}{A_i}) & , A_i < 0 \end{cases} \tag{11.3}$$

Spectroscopic features that are, in average, abnormal in large peritumoral volumes are associated with larger $D_{abn,i}$ values.

The *MRS-FSD* was fitted to the median curves of the 14 different features, i.e. the curves described by the median feature values in function of the distance to STV. This was motivated by the fact that the number of voxels decreases considerably as the distance to the tumor increases. Moreover, the STV voxels (d=0) have often skewed distributions. Finally, the use of the median curves instead of the voxels separately improves the robustness of the results to the presence of outliers.

In order to test the stability of the results, the fitting was performed 6-times, excluding every time the data from one-sixth of the patients. This was done on par with the 6-fold cross-validation scheme used to test the predictions of the distance to the STV described in the following section. The mean and standard deviation of the fitting results were determined.

11.3.7 Expected Distance to Tumor (EDT) Maps

To test the ability to identify metabolic patterns associated with different distances to the STV, a regression forest [Breiman 2001] was trained to estimate the distance to solid tumor based on all of the 14 features previously described. For comparison, the distance to the STV was also predicted based on individual features using the corresponding inverted MRS-FSD models. In both cases, the training followed a 6-fold Cross-Validation (CV) scheme, where in each iteration the data from one-sixth of the patients were excluded from the training and used to test the trained model.

The regression forests [Breiman 2001] (R, randomForest package [Liaw 2002]) were trained using 500 trees and standard parameters (minimum size of terminal nodes equal to 5, and number features per node equal to one-third of the total number of features [Hastie 2001], i.e., $14/3 = 4.667 \approx 5$). The trained regressors were used to evaluate which features carry the most information to estimate EDT values. Two different measures of feature importance were analyzed: increase in *Mean Squared Error* (MSE) and *residual sum of squares*. The first one is calculated as the average increase in error as a consequence of permuting the values of a given

feature between different Out-Of-Bag (OOB) examples, a group of examples that are excluded from the training in each tree as a result of *bagging* [Breiman 1994]. The greater the error increase, the greater the importance of the permuted feature for the predictions. Regarding the second measure, the residual sum of squares, this measures the ability of a given feature to separate the examples into different groups with different properties. Features that, during training, have a higher contribution to the separation of the training examples, are considered more important. More details on these feature importance measures that are characteristic of regression forests can be found in the original paper from Breiman [Breiman 2001].

In order to allow the regression forest models to distinguish the different tumor compartments, virtual distances were assigned to the voxels of these classes before training. A distance of -2 mm was used for necrosis, -1 mm for E-Tumor and 0 for NE-Tumor.

Regarding the estimation of the distance to STV based on the value of a feature i and the corresponding MRS-FSD model, this were determined as:

$$d(x_i) = \begin{cases} -\lambda_i \cdot \log(\frac{x_i - K_i}{A_i}) & , \frac{x_i - K_i}{A_i} \geq \varepsilon \\ -\lambda_i \cdot \log(\varepsilon) & , \frac{x_i - K_i}{A_i} < \varepsilon \end{cases} \tag{11.4}$$

where ε is an arbitrarily small number that, in this work, was defined as 0.001. This prevents problems with close-to-0 and negative arguments of the logarithm function and sets the maximum expected distance that can be predicted using the corresponding model.

11.4 Results

11.4.1 Spectra per MRI Segmentation Class

Figure 11.1 shows the different spectra for each MRI segmentation class. The results show that the spectra vary considerably for each MRI class, especially for the tumor classes. This is caused not only by the differences between the information provided by MRS and sMRI [Lopez 2017, Pirzkall 2004], but also by the partial volume effects derived from the low resolution of the acquired MRSI data. Despite the variability, the grouping of the spectra per MRI-segmentation class highlights several distinctive characteristics of the different tissues. As expected, the highest levels of Lip and Lac, and lowest levels of NAA and Cr are normally found in necrosis. Enhancing and Non-Enhancing Tumor have very similar spectra, reflecting that contrast leakage caused by blood-brain-barrier disruption does not necessarily translate into differences in metabolic features. Interestingly, in Edema we find the highest levels of Glx (see difference spectrum between 3.7 and 3.8 ppm, and between 2.1 and 2.4 ppm).

Regarding the CSF spectra, pure CSF spectra are expected to contain no metabolites, which conflicts with the results of Figure 11.1. However, as described in the methods section, each spectrum was assigned to the MRI-segmentation class of the *majority* of the corresponding image voxels. This means that, due to this strategy and the low resolution of the MRSI protocol used, most of the spectroscopic voxels are not "pure" CSF voxels, which explains the existence of visible metabolites in the "CSF" spectra presented here. Moreover, since all spectra are normalized, the height of the peaks in CSF is identical to the one seen in GM or WM.

11.4.2 Spectra per distance to the STV

Figure 11.2 shows the magnitude spectra depending on the distance to the STV. In this plots, a distance of 0 mm corresponds to all spectra of the classes of the STV (E-Tumor, NE-Tumor,

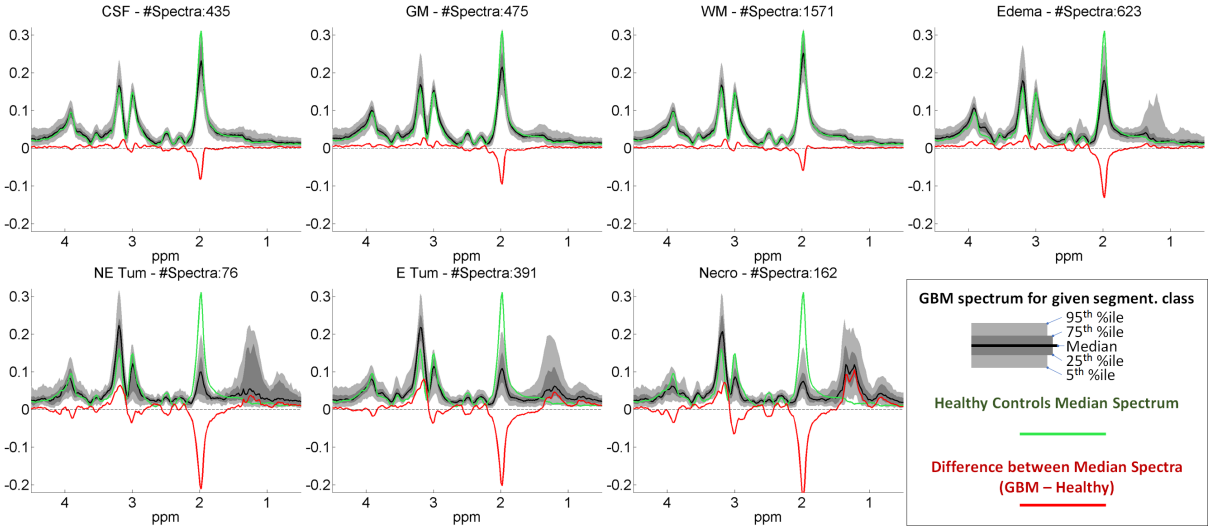

Figure 11.1: Spectra per MRI-Segmentation class. The black line shows the median magnitude spectra for each class and the shades of the gray show the corresponding 5^{th}, 25^{th}, 75^{th} and 95^{th} percentiles. The green line shows the median magnitude spectra of data collected from healthy controls (includes GM, WM and CSF), and the red line the difference between the two median magnitude spectra. All spectra were normalized using L_1-normalization. The normal brain tissue pattern seen for the CSF class is a consequence of the partial volume effects and the normalization of the spectra.

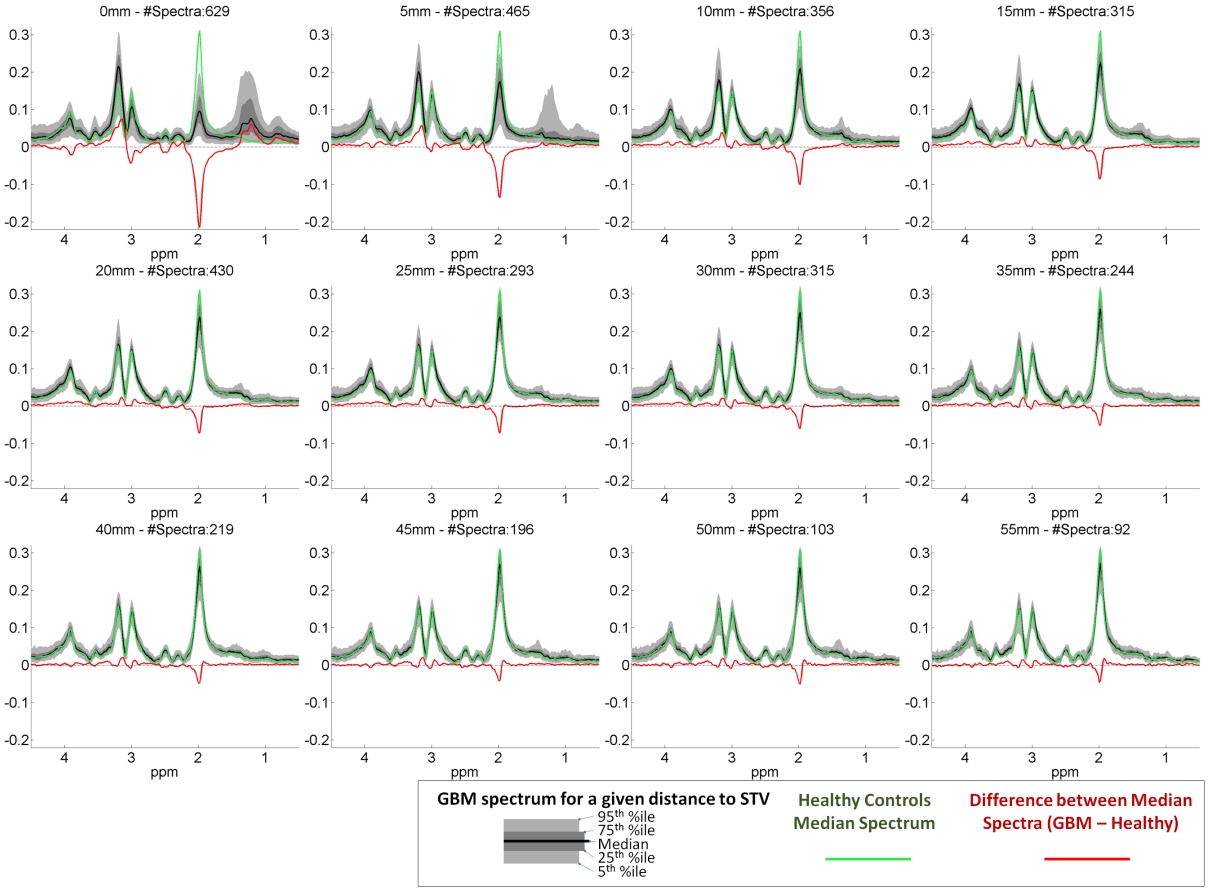

Figure 11.2: Spectra for different distances to the STV. The black line shows the median magnitude spectra for each distance and the shades of the gray show the corresponding 5^{th}, 25^{th}, 75^{th} and 95^{th} percentiles. The green line shows the median magnitude spectra of data collected from healthy controls (includes GM, WM and CSF), and the red line the difference between the two median magnitude spectra. All spectra were normalized using L_1-normalization.

Necrosis). By visual inspection, it is difficult to detect changes in distances greater than 30/35 mm. From distances between 5 and 30 mm, the most noticeable effects are the increase in NAA and decrease in Cho as the distance to tumor increases. For distances of 5 mm, some spectra contain elevated levels of lipids, which should come as result of partial volume effects or errors in MRI segmentation. A closer look at the spectra shows that Glx levels are visibly elevated in regions with distances between 5 and 15 mm from the tumor.

11.4.3 MRS features as a function of the distance to the STV and Image Segmentation Class

The values of the six *%metabolite* features as a function of the distance (blue) and MRI segmentation class (green) are presented in Figure 11.3. The reference values of the 5^{th}, 25^{th}, 50^{th}, 75^{th} and 95^{th} percentiles of each feature in healthy volunteers are shown in red for reference. For sake of simplicity, the plots corresponding to the remaining metabolite ratios are included in the Supporting Information (Figures 11.10 and 11.11).

The data shows that %NAA and %Cr levels are reduced in the brain of GBM patients, even in regions that are located very far from *STV*. As expected, both features get further reduced, as the distance to *STV* decreases. In contrast, the relative quantities of the remaining metabolites increase from normal brain tissue to tumor. The feature %Glx shows a unique characteristic: the maximum value is observed in the peripheral zone of the tumor, in edema and for distances between 5 and 10 mm from the STV.

11.4.4 MRS-FSD model

Table 11.1 shows the results obtained by fitting the MRS-FSD model described (Eq. 11.1) to the rescaled data (see Eq. 11.2). Figure 11.4 shows a plot of the spectral feature normalization lengths and amplitudes for all ratios except %Glx, for which the normalization length is considerably larger when compared to the remaining metabolites. The introduced model is able to well describe the relation between the feature values and the distance to STV, as shown by the high R^2 values, with the exception of %Cho. For this feature, the differences in value between tumor and healthy brain tissue are close to the level of the noise and, therefore, the model fails to explain a considerable part of the variance of the data. The plots showing the fitting results are included in the Supporting Information (Figures 11.12 to 11.25).

Regarding Glx ratios, Glx/NAA seems to be more sensitive than Glx/Cr to detect the effects of tumor presence, showing a larger normalization length and amplitude. This result can be explained by the fact that, compared to Cr, NAA is reduced in a broader volume that extends farther from the STV, as shown by the larger absolute value of the amplitude and larger normalization length. Given that the increase in Glx has been associated with neuronal loss and demyelination [Ricci 2007], it is expected that the ratio with NAA increases the sensitivity to detect deterioration and loss of neurons.

The Lip and Lac ratios showed the smallest normalization lengths, reflecting the fact that these features are significantly altered only in a very confined region, namely in necrotic tissue.

Besides Cho/NAA, whose sensitivity to detect peritumoral changes has been extensively discussed in the literature, two other features stand out as a result of their large abnormality distances: %NAA and Glx/NAA.

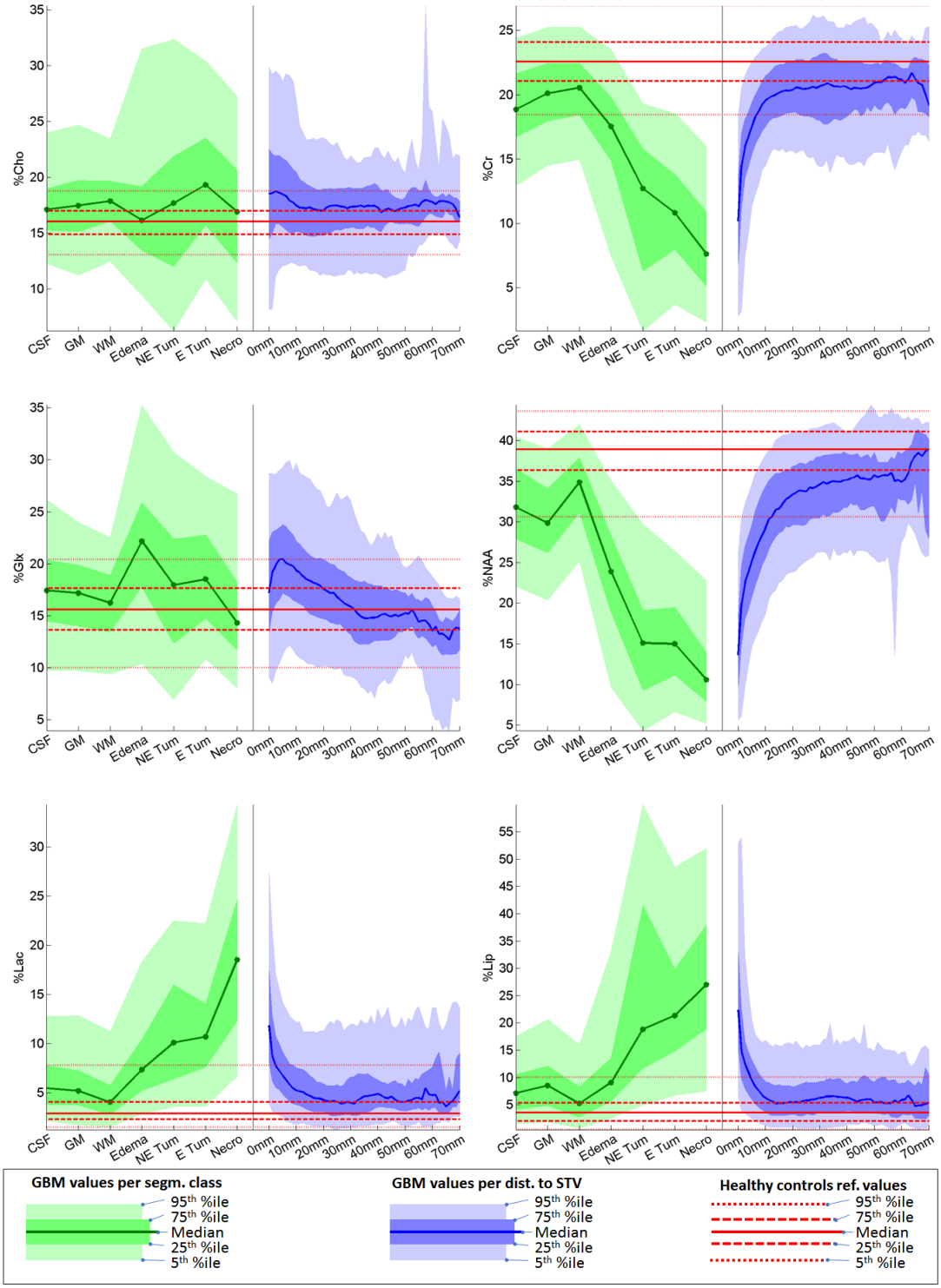

Figure 11.3: Value of several *%metabolite* features as a function of the sMRI segmentation class (green), and the distance to *STV* (blue). For each value of the horizontal axis, the 5^{th}, 25^{th}, 50^{th}, 75^{th} and 95^{th} percentiles are shown as depicted in the legend. For each feature, the horizontal level lines in red mark the 5^{th}, 25^{th}, 50^{th}, 75^{th} and 95^{th} percentiles in healthy volunteers (includes GM, WM and CSF).

Table 11.1: Fitting results of the *MRS-FSD* model for different features. The model was fitted to the median curve of the feature values in function of the distance to STV. Before curve fitting, the data was rescaled as described in expression 11.2. For each feature, model fitting was performed six times, excluding every time the data from one-sixth of the patients. The mean and standard error of the mean (SEM) of the fitting results are shown for the coefficient of determination (r^2), model parameters (K, amplitude A and normalization length λ) and abnormality distance (D_{abn}). Values of λ and D_{abn} are given in mm.

Feature	mean R^2	SEM R^2	mean K	SEM K	mean A	SEM A	mean λ	SEM λ	mean D_{abn}	SEM D_{abn}
%Cho	0.470	0.073	0.746	0.008	0.290	0.041	6.963	1.273	0.354	0.730
%Cr	0.943	0.005	0.274	0.014	-1.121	0.018	4.623	0.305	6.479	0.277
%NAA	0.946	0.006	0.406	0.027	-1.484	0.018	9.571	0.802	12.308	0.600
%Glx	0.886	0.008	0.003	0.100	0.996	0.097	60.884	9.453	-0.423	1.140
%Lip	0.925	0.015	0.562	0.011	1.650	0.053	3.043	0.162	4.008	0.119
%Lac	0.848	0.060	0.465	0.017	1.064	0.023	4.393	0.295	3.004	0.179
Cho/NAA	0.967	0.003	0.725	0.020	3.357	0.058	4.230	0.251	10.602	0.558
Cho/Cr	0.975	0.001	0.903	0.012	2.700	0.075	3.162	0.083	10.583	0.265
Glx/NAA	0.959	0.001	0.500	0.017	2.379	0.021	8.706	0.238	13.577	0.235
Glx/Cr	0.948	0.006	0.470	0.014	1.389	0.015	9.409	0.373	9.043	0.263
Lip/Cr	0.959	0.004	0.631	0.009	4.250	0.188	1.915	0.084	4.660	0.150
Lac/Cr	0.930	0.013	0.540	0.011	3.228	0.072	2.183	0.087	4.252	0.161
Lip/NAA	0.961	0.005	0.515	0.012	4.293	0.175	2.064	0.101	4.482	0.158
Lac/NAA	0.930	0.012	0.451	0.017	3.252	0.095	2.547	0.113	4.529	0.205

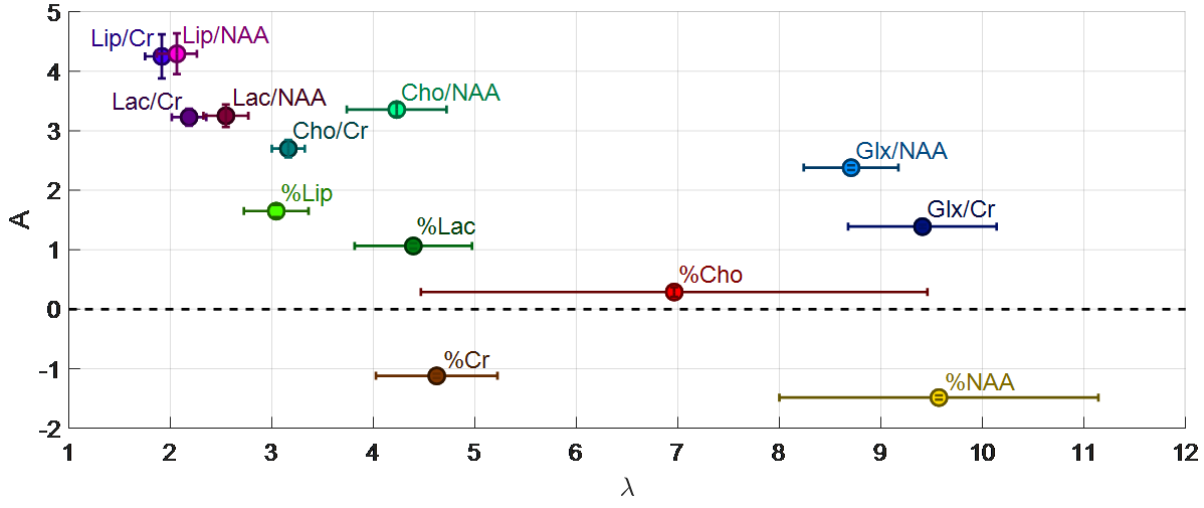

Figure 11.4: Plot of the different amplitudes (A) and normalization lengths (λ) for each feature. The plot shows the mean values ± 1.96 times the standard error of the mean of the fitting results shown in Table 11.1. The results of %Glx are not shown in this plot, given that this feature shows a considerably larger normalization length relative to the other ratios.

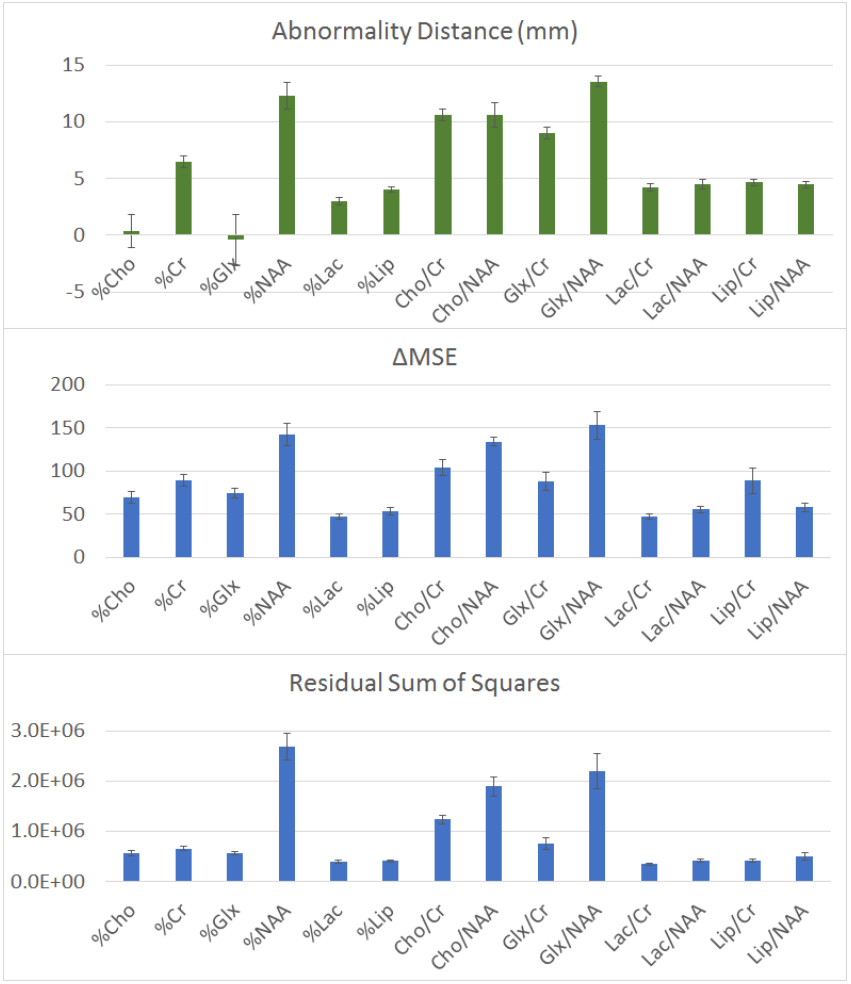

Figure 11.5: Abnormality distances and relative importance for the prediction of the distance to STV of the different MRS features. The abnormality distances (green) were calculated based on the MRS-FSD models for each feature. The two feature importance measures (blue) were determined by the RF-models trained to estimate the distance to the tumor. The bar plots show, for each feature, the mean values ± 1.96 times the standard error of the mean of the 6-fold CV results. For a better explanation of the meaning of the feature importance measures please see the Section 11.3.7.

11.4.5 Expected Distance to Tumor (EDT)

The feature importance results associated with the regression forest models, as well as the D_{abn} values for each feature, are shown in Figure 11.5. As it can be seen, the abnormality distances are highly correlated with feature importance for the prediction of the distance to STV. The Pearson correlation coefficient between the average values of D_{abn} and ΔMSE was 0.836 (p=0.0002), and between the average values of D_{abn} and RSS 0.833 (p=0.0002).

The results show that the four features that contain the most information to estimate the distance to the tumor are %NAA, Glx/NAA, Cho/NAA and Cho/Cr. From the Lip and Lac related features, Lip/Cr seems to be the most important. Since the different tumor compartments were labeled with different virtual distances (-2 for Necrosis, -1 for E-Tumor and 0 for NE-Tumor), it is expected that the importance of the Lip/Cr ratio is related to its role in identifying necrotic tissue.

The comparison between the estimated and measured distances to STV for the RF model and for the inverted MRS-FSD models of 5 of the most important features, are presented in Figure 11.6. Compared to the single-feature models, the predictions made by the RF-model are more accurate and show a smaller variance. However, the predictions of the RF-model show a positive bias for regions close to the STV, and a negative bias in regions far from the solid tumor. Interestingly, the median curve of the RF-model is approximately monotonic. Due to the superiority of the RF-model, this was the model selected to generate the EDT values used in the following results of this paper.

Figure 11.7 shows the values of different features per EDT and Figure 11.8 the characteristic spectra associated with different EDT values. These results show the relations between feature values and distance to the STV that were learned by the RF model. The different plots of Figure 11.8 show that, for each EDT value, there is a specific spectroscopic pattern associated with it. Compared with the results of Figure 11.2, which show the spectra for different distances to STV, the grouping of the spectra per EDT value shows a better ordering of the spectra in terms of malignancy. More specifically, the ordering per EDT value shows more pronounced malignant features, i.e. higher Cho/NAA and Cho/Cr, for shorter (expected) distances and less malignant features, i.e. spectra identical to normal brain spectra, for spectra with large (expected) distances to the tumor.

Finally, Figure 11.9 shows the sMRI segmentation overlaid with the isolines of the corresponding EDT maps. In general, the EDT maps show a good agreement with the sMRI segmentation. However, in several of these maps, regions identified as edema or even normal brain tissue have small EDT values, which, potentially, may indicate regions of tumor infiltration.

11.5 Discussion

11.5.1 MRS data shows significant metabolic alterations far from solid tumor visible in MRI

The results of Figure 11.3 show that several of the metabolite ratios are significantly altered in regions outside the STV. Previous studies have shown that MRS is not only able to detect changes being solid tumor visible MRI but also beyond any T_2 abnormalites [Cordova 2016, Guo 2012, Stadlbauer 2004, Ganslandt 2005], a characteristic that has been explored to improve tumor delineation. It is expected that the main contribution to the peritumoral changes seen in Figure

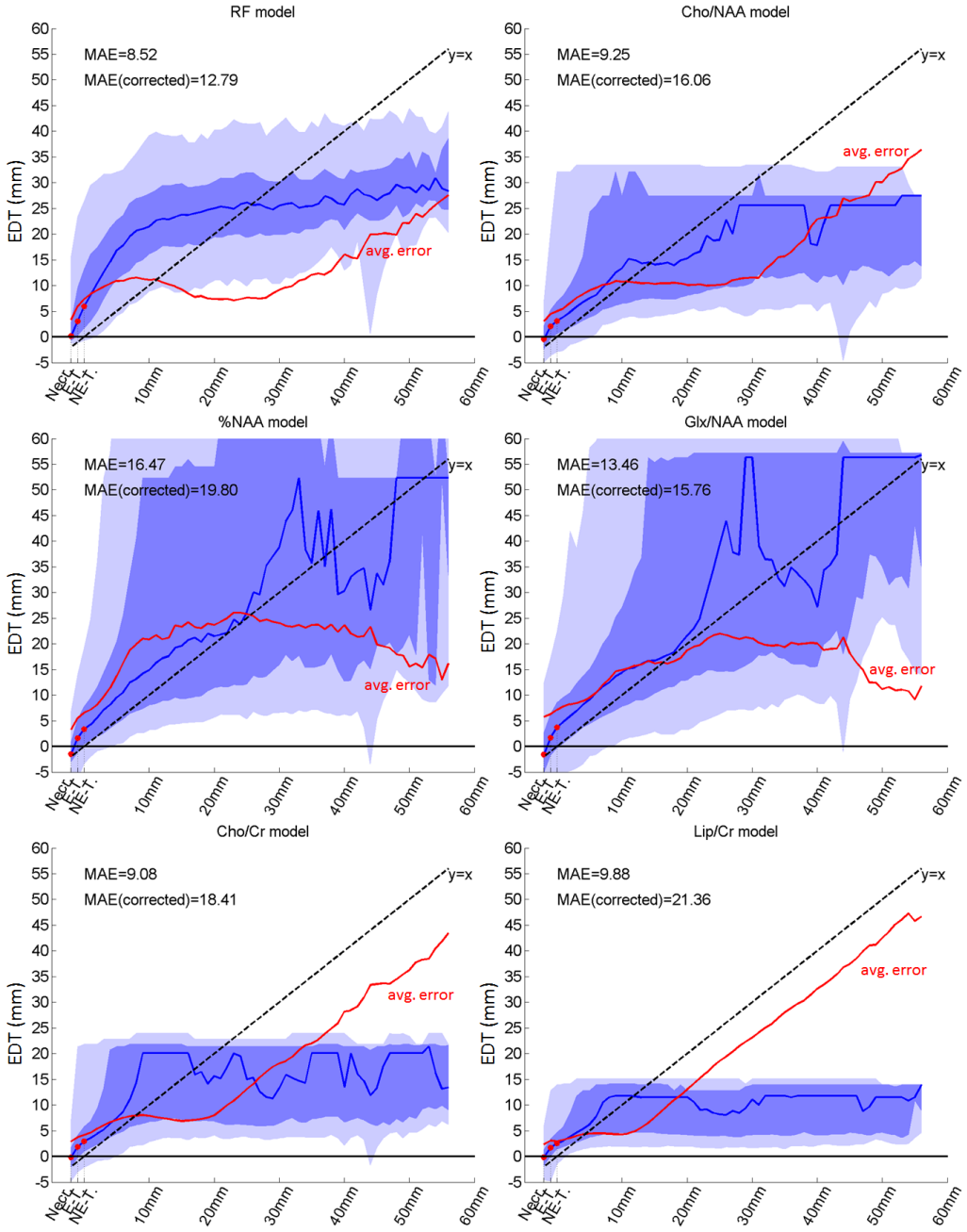

Figure 11.6: Comparison between the predictions of the distance to the STV made using different models. The plots show the predicted distances as a function of the measured distances to the tumor. In each plot, the blue line and the blue shades show the median and the 5^{th}, 25^{th}, 75^{th} and 95^{th} percentiles of the predicted distances as a function of the measured distances to the tumor, similarly to the shades seen in the plots of Figure 11.3. The red line shows the mean absolute error in function of the distance to STV. In each plot, the mean absolute error (MAE) for all voxels and corrected MAE are shown. The corrected MAE corrects for the differences in the number of voxels of different distances and corresponds to the average value of MAE for each distance, i.e. the average value of the red curve. The virtual distances used to distinguish the different tumor compartments are highlighted in each plot (-2 for Necrosis, -1 for E-Tumor and 0 for NE-Tumor). All results produced using 6-fold CV.

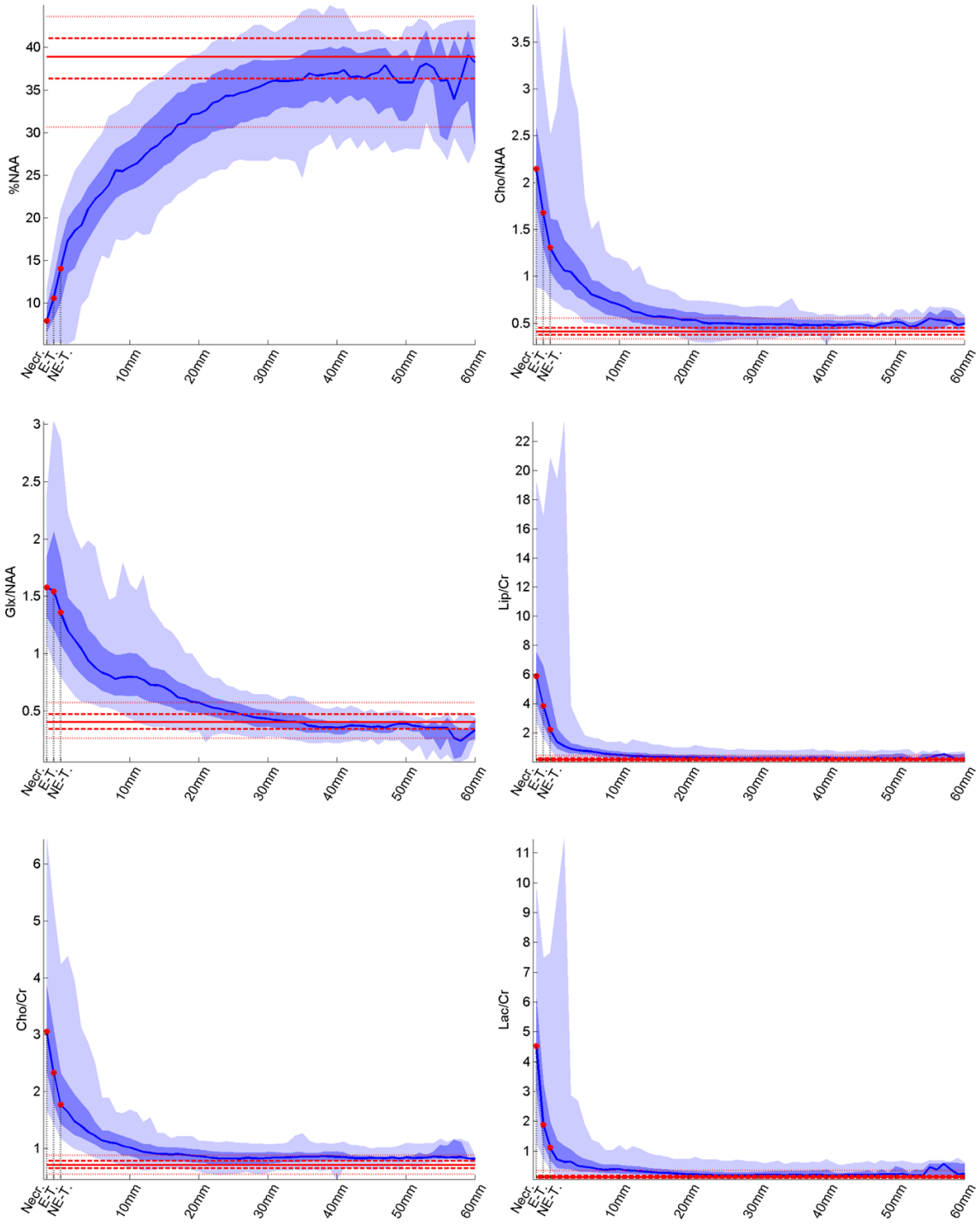

Figure 11.7: Feature values in function of the EDT values predicted using 6-fold CV. For each value of the horizontal axis, the 5^{th}, 25^{th}, 50^{th}, 75^{th} and 95^{th} percentiles are shown (blue shades), as in Figure 11.3. For each feature, the horizontal level lines in red mark the 5^{th}, 25^{th}, 50^{th}, 75^{th} and 95^{th} percentiles in healthy volunteers (includes GM, WM and CSF). The negative and zero EDT values correspond to Necrosis (-2 mm), E-Tumor (-1 mm) and NE-Tumor (0 mm), as it is highlighted in every plot.

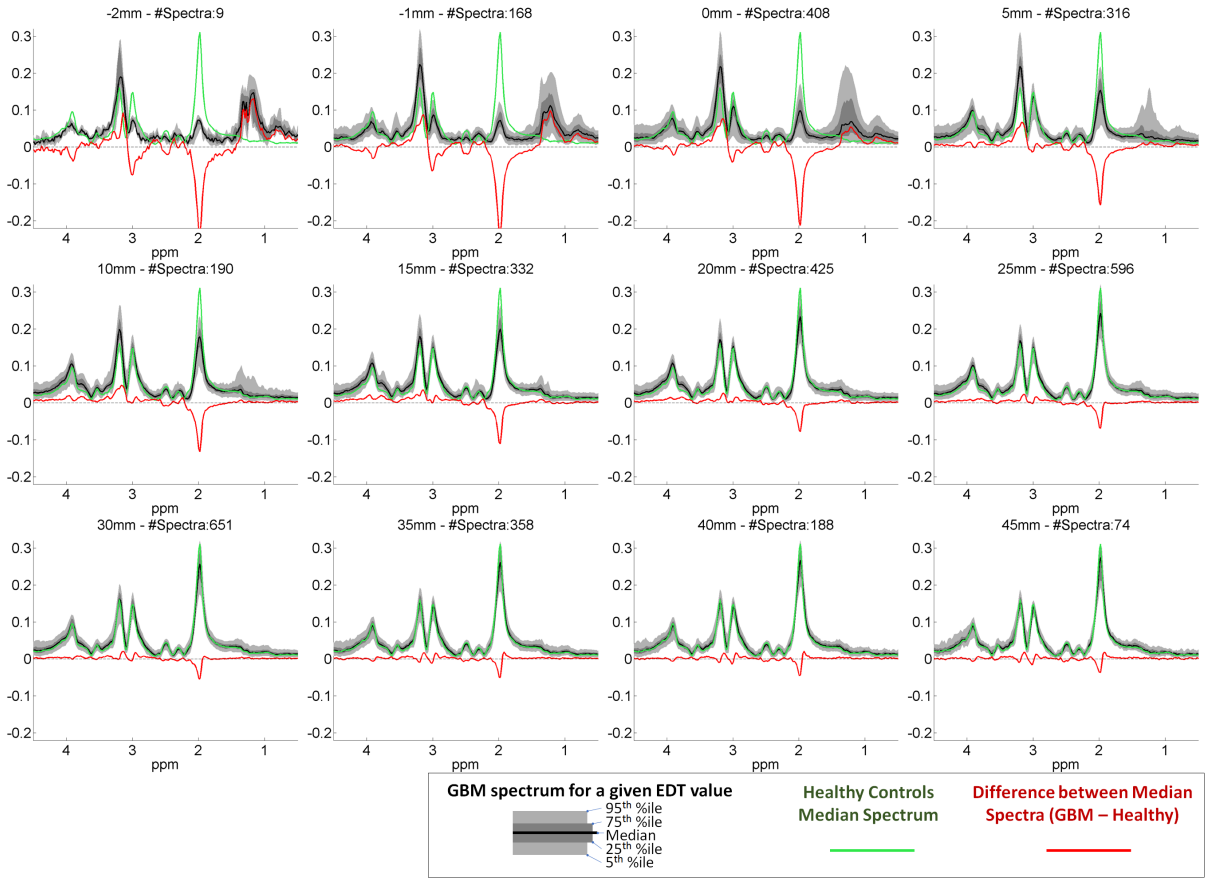

Figure 11.8: Spectra for different EDT values predicted using 6-fold CV. EDT values were rounded to the closest distance value shown in these plots. Spectra with EDT values equal or bigger than 47.5mm were not included. The negative and zero EDT values correspond to Necrosis (-2 mm), E-Tumor (-1 mm) and NE-Tumor (0 mm). The black line shows the median magnitude spectra for each class and the shades of the gray show the corresponding 5^{th}, 25^{th}, 75^{th} and 95^{th} percentiles. The green line shows the median magnitude spectra of data collected from healthy controls (includes GM, WM and CSF), and the red line the difference between the two median magnitude spectra. All spectra were normalized using L_1-normalization. No spatial interpolation was performed.

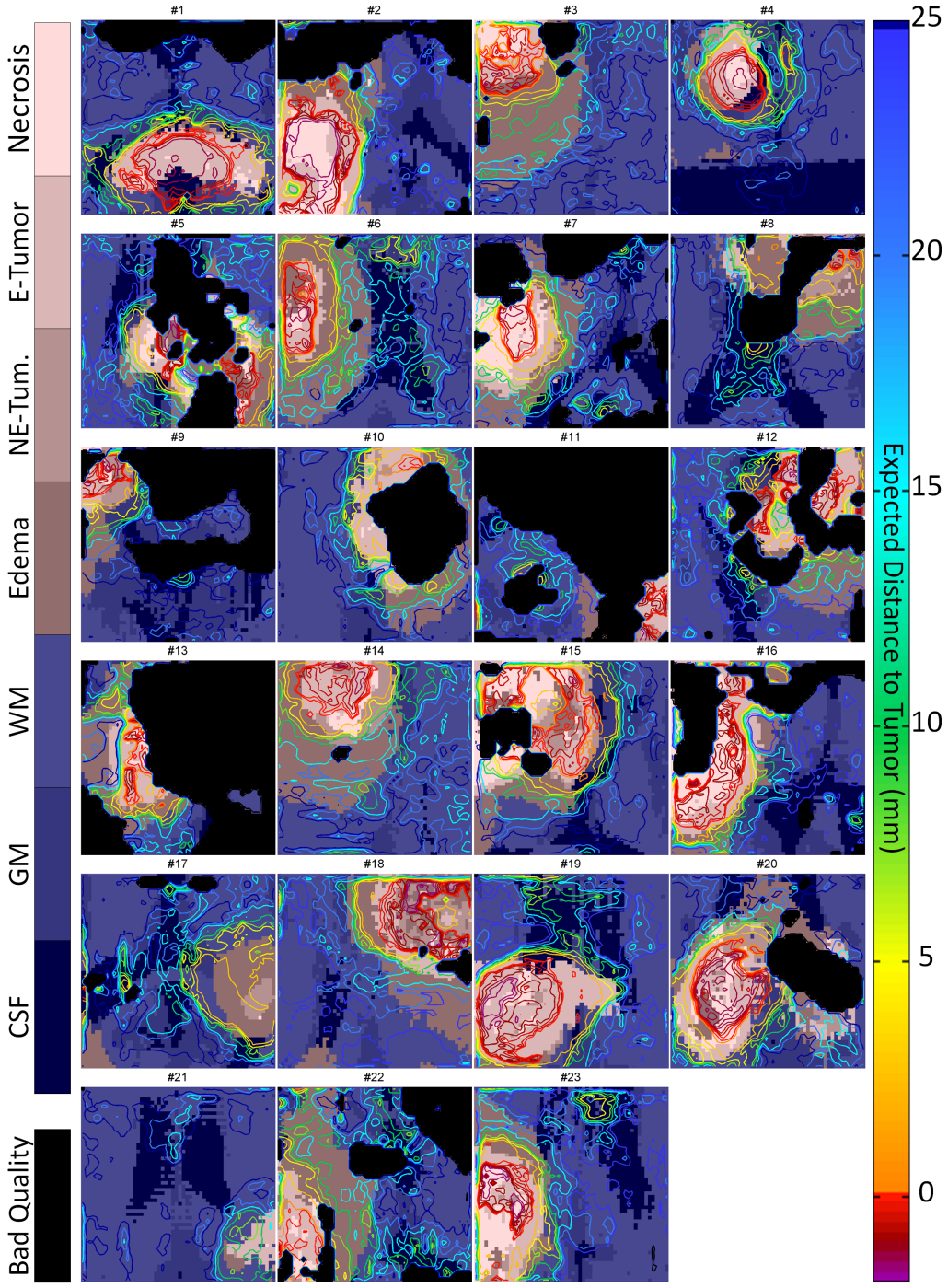

Figure 11.9: Isoline maps of the Expected Distance to Tumor (EDT), determined based on the spectroscopic information from each voxel, overlaid over the corresponding sMRI segmentation. The EDT maps were calculated using the described RF-model and following a 6-fold CV scheme, i.e. the maps from each row where calculated using models trained based only on the data from the other rows. The color code used for the background image (sMRI segmentation) is shown on the left side. All regions for which the spectra had insufficient quality are shown in black. The color scale of the EDT isolines is shown on the right.

11.3 originate from phenomena of clinical interest, such as tumor infiltration, edema, and inflammatory response. This is supported by the mentioned results from the previous literature, as well as by the structure seen in the peritumoral region that is made visible by the EDT maps of Figure 11.9. Despite this, the low resolution of the acquired data means that effects caused by signal bleeding [Pedrosa de Barros 2017b, Kreis 2004, Bovée 1998], which causes the voxels near the tumor to have altered spectra regardless of the presence of any tumor-related effect, cannot be ignored. The severity of the signal bleeding depends not only on the native resolution of each MRSI examination but also on other factors such as k-space filtering and on the shape of the source of the signal, in this case, the tumor. The impact that signal bleeding can have on MRSI data can be assessed by measuring and analyzing the point-spread-function (PSF), as described for instance in [Bovée 1998], or simply performing phantom studies to evaluate the accuracy of tumor delineation as described in [Heikal 2008]. These two aspects should be included in further studies. Besides studies to evaluate the impact of the low resolution in the analysis of the relation between the distance to tumor and MRSI features, it is essential to perform a similar study using higher resolution MRSI data acquired, for instance, with EPSI [Posse 1995, Sabati 2015]. For an even finer resolution, upscaling of MRSI can be performed based on the structure found in high-resolution MRI images, as shown in [Jain 2017]. However, in this case, it would be necessary to evaluate the use of the mentioned super-resolution method for brain tumor data.

11.5.2 The MRS-FSD model, a tool to analyze the spatial properties of MRS features

The MRS-FSD model introduced in this paper describes the main aspects of the relation between each MRS-feature and the distance to STV. More precisely, the model characterizes each feature in terms of (**1.**) the magnitude and sign of the differences between healthy brain tissue and solid tumor, and (**2.**) the spatial rate of the change. This allows to identify different groups of features: features that are highly altered locally in solid tumor, such as those associated with tumor necrosis like Lip/Cr and Lip/NAA; features that are reduced in solid tumor, such as %Cr and %NAA; and features showing abnormal values in a broad peritumoral region, such as Glx/NAA and %NAA. For the detection of tumor-related changes, the latter group is obviously the most interesting. To help the interpretation of the model parameters, the abnormality distance D_{abn} was introduced. As it was shown, D_{abn} can be used to identify the features that carry the most information for determining the distance to STV. These results, show that the MRS-FSD model is a simple tool that may be used to assist the analysis of the spatial properties of different MRS features in relation to the distance to the solid tumor. Nevertheless, it should be highlighted that the simplicity of the model does not allow to completely describe features such as %Glx, for which the maximum values occur neither in STV or healthy brain tissue but in between these two regions.

11.5.3 Glx/NAA and %NAA may allow detecting early signs of tumor infiltration

The results shown in Figure 11.5, show that %NAA and Glx/NAA are the metabolite ratios with the largest D_{abn}. These features are also the ones that contain the most information for predicting the distance to the STV, followed by Cho/NAA, a metabolite ratio that is known to be associated with tumor infiltration [Cordova 2016]. The ability of Cho/NAA to identify tumor infiltration is justified by the combination of two opposed effects [Oz 2014, Govindaraju 2000, Mountford 2010]: on the one hand, the high membrane turn-over seen in tumor cells leads

to high values of (free) Cho and, on the other hand, the neuronal loss/dysfunction as a result of the tumor causes a decrease in NAA. Despite elevated Cho levels being associated with tumor metabolism, the results of Cordova et al. [Cordova 2016] show that Cho/Cr is not a good predictor for tumor infiltration, which may suggest that NAA is the metabolite that mainly explains the correlation between Cho/NAA and tumor cellularity. This is supported by the results presented in this paper, showing that %Cho (not to be confounded with the concentration of Cho) is relatively constant regardless of the distance to the tumor or the tumor compartment, in contrast with %NAA, the "%metabolite" feature that shows the biggest differences between normal brain tissue and solid tumor. These results further indicate that %NAA might be a good marker for detecting infiltration. Regarding the ratio Glx/NAA, high values of this ratio can occur as a result of an increase in Glx, which has been associated with inflammation, neuronal loss, demyelination and tumor proliferation [Ricci 2007, Takano 2001, DeBerardinis 2010], combined with a decrease in NAA. This combination of effects suggests that, similarly to %NAA, Glx/NAA might also be an interesting metabolite for the detection of tumor infiltration. Future studies are needed to evaluate the value of these two metabolite ratios for the prediction of tumor infiltration.

11.5.4 EDT maps facilitate clinical interpretation of MRSI data

The RF-model trained to predict the distance to the STV based on the different metabolite ratios translates each spectrum into a single EDT value. Moreover, this model was trained to distinguish different tumor compartments, since Necrosis, E-Tumor and NE-Tumor were labeled with different virtual distances (0 for NE-Tumor, -1 for E-Tumor and -2 for Necrosis). The results shown in Figures 11.8 and 11.7 seem to indicate a clear association between metabolic patterns and EDT values. This characteristic suggests that EDT maps may be regarded as a novel of type of nosologic image, allowing an easier clinical interpretation of GBM MRSI data. Moreover, compared to blind source separation methods [Sajda 2004, Du 2004, Du 2008, Su 2008, Ortega-Martorell 2012a, Li 2013b, Li 2013a, Raschke 2015, Mocioiu 2016, Pedrosa de Barros 2015], no assumptions on the number of basis spectra that need to be considered are required, which may simplify its implementation in clinical practice. Despite this, it would be important to study the correspondence between EDT values and the tissue contributions identified by blind source separation methods.

11.5.5 EDT maps may improve detection of infiltration

As described in the previous paragraph, EDT models convert MR spectra into a single "expected distance" that is hypothesized to have a correspondence to a stage in tumor growth. Compared to Cho/NAA thresholding, the RF-based model uses multiple features and is able to learn complex relations between them. Consequently, under the assumption that the probability of tumor infiltration decreases as the distance to STV increases, it is possible that RF-based models may provide a better detection of tumor infiltration than Cho/NAA maps alone. In the case of EDT maps, tumor infiltration is expected to be represented by regions of low EDT values that reflect that these regions share some of the properties seen in regions close to the STV, where infiltration is most likely. An encouraging fact that supports the idea that EDT may help identify infiltration is that, as we see in Figure 11.6, the RF-model shows that for distances to the STV greater than 20 mm, there is less information for determining EDT values. This suggests that spectra with distances greater than 20 mm are identical to normal brain spectra. These results match the findings described by Pallud et al. [Pallud 2010], that describe that for diffuse

low-grade gliomas MIB-1-positive cells are not normally found in samples collected at distances greater than 20 mm from MRI-defined abnormalities. The relation between EDT maps and tumor infiltration should be validated in future studies. Moreover, it could be interesting to compare the performance for detecting tumor infiltration between the EDT method and blind source separation methods.

11.5.6 EDT maps can be easily implemented in clinical practice

The data required to train a new model for generating EDT maps consists of a set of processed MRSI examinations and the corresponding sMRI segmentations. In this work, both the processing of MRSI data and the segmentation of sMRI were performed fully automatically using, respectively, SpectrIm and BraTumIA. Though these two software applications are currently only available for research, they demonstrate that it is feasible to fully automatically generate the training data based on raw examinations. Once the model is trained, the quantification results of each MRSI examination can be converted into EDT maps almost instantaneously.

The automation of the process of training new EDT models does not only simplifies its implementation but also ensures the reproducibility of the results between centers. In case automatic tumor segmentation is not available, semi-automatic methods may also enable reproducible identification of the STV in sMRI [Huber 2017].

Finally, another factor facilitating the introduction of the EDT maps in clinical practice is the easy interpretation of the results: an EDT value of 10 mm, for instance, can be easily interpreted as a metabolic pattern that is likely to be found at a distance of 10 mm from the solid tumor border.

Despite these advantages, the results shown in this article can only be regarded as a proof of concept, since clinical validation of the method is currently missing.

11.6 Conclusion

The results of the presented study confirmed that most ^{1}H-MRS visible metabolites are altered beyond the STV. Moreover, using the MRS-FSD model, it was possible to compare the spatial properties of several of the most commonly used metabolite ratios for the characterization of brain tumors. The analysis suggests that besides Cho/NAA, features such as %NAA and Glx/NAA, which are often significantly altered far from the tumor, may allow detecting early signs of tumor infiltration. Finally, the relation between the distance to the STV and the values of the different metabolite ratios was explored to produce a novel type of map: the EDT maps. This approach converts the information concerning multiple metabolites into a single number, whose value may have a correspondence with the different tissues seen at different stages of tumor growth. EDT maps simplify the interpretation of MRSI data of brain tumor patients and may help to identify early signs of tumor infiltration. Further studies should explore this new concept with higher resolution MRSI data and evaluate its potential application for the diagnosis, follow-up, neurosurgical treatment, and radiotherapy planning of glioma patients.

Acknowledgements

This work was funded by the EU Marie Curie FP7-PEOPLE-2012-ITN project TRANSACT (PITN-GA-2012-316679) and the Swiss National Science Foundation (project number 140958).

11.7 Supporting Information

11.7.1 Metabolite ratios per sMRI segmentation class and per distance to STV

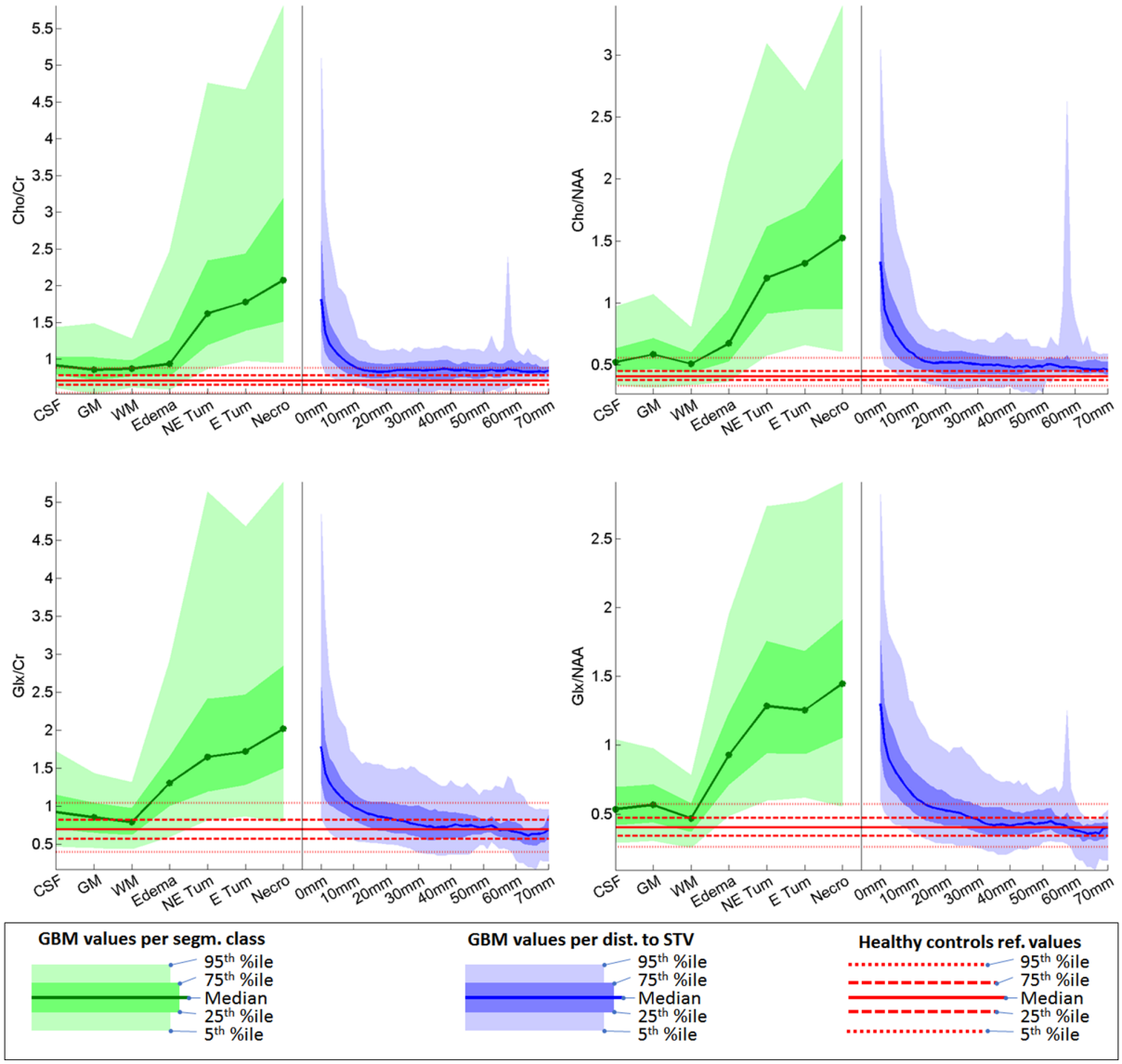

Figure 11.10: Value of several features as a function of the sMRI segmentation class (green), and the distance to STV (blue). For each value of the horizontal axis, the 5^{th}, 25^{th}, 50^{th}, 75^{th} and 95^{th} percentiles are shown as depicted in the legend. For each feature, the horizontal level lines in red mark the 5^{th}, 25^{th}, 50^{th}, 75^{th} and 95^{th} percentiles in healthy volunteers (includes GM, WM and CSF).

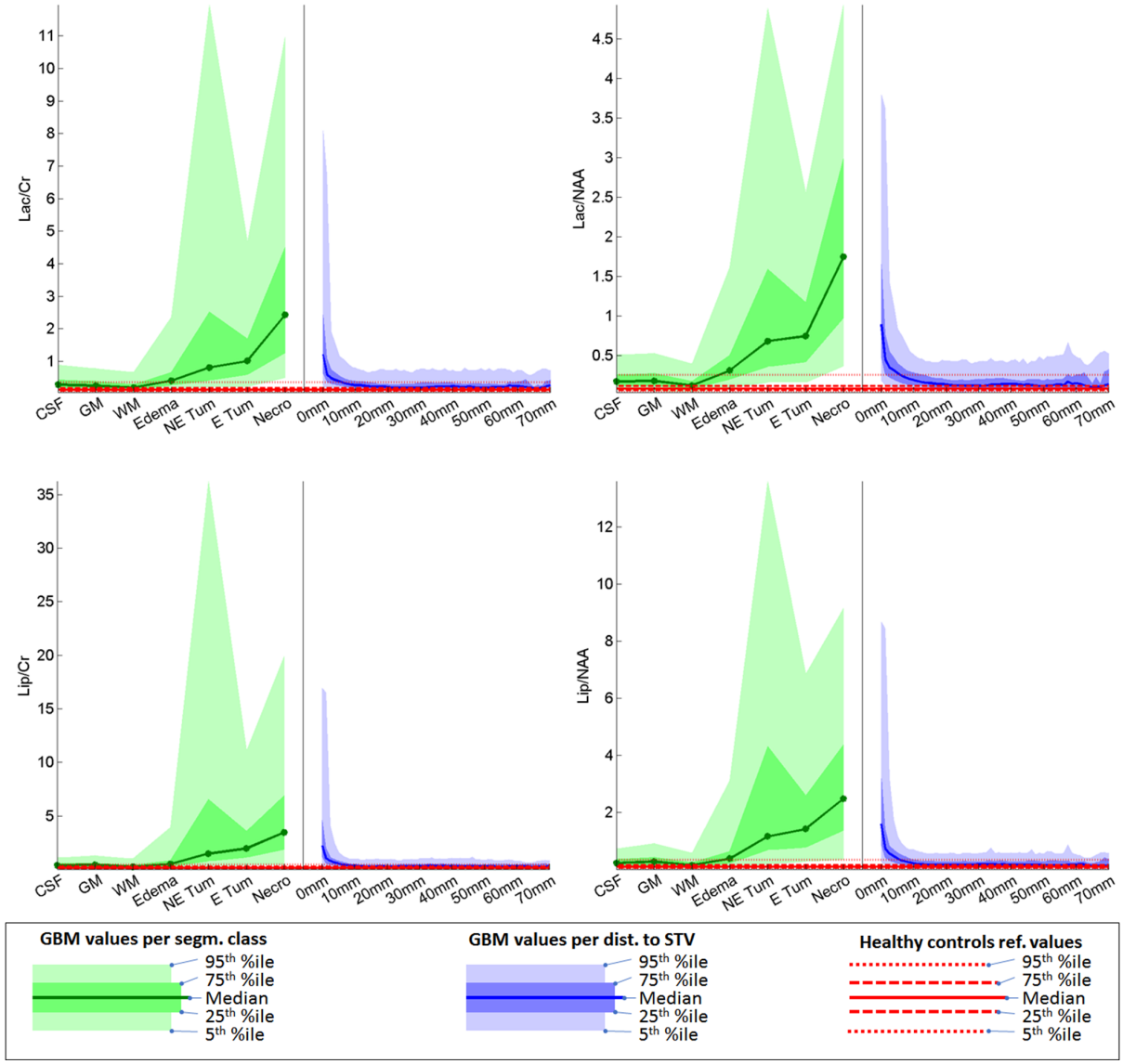

Figure 11.11: Value of several features as a function of the sMRI segmentation class (green), and the distance to STV (blue). For each value of the horizontal axis, the 5^{th}, 25^{th}, 50^{th}, 75^{th} and 95^{th} percentiles are shown as depicted in the legend. For each feature, the horizontal level lines in red mark the 5^{th}, 25^{th}, 50^{th}, 75^{th} and 95^{th} percentiles in healthy volunteers (includes GM, WM and CSF).

11.7.2 MRS-FSD Fitting Results

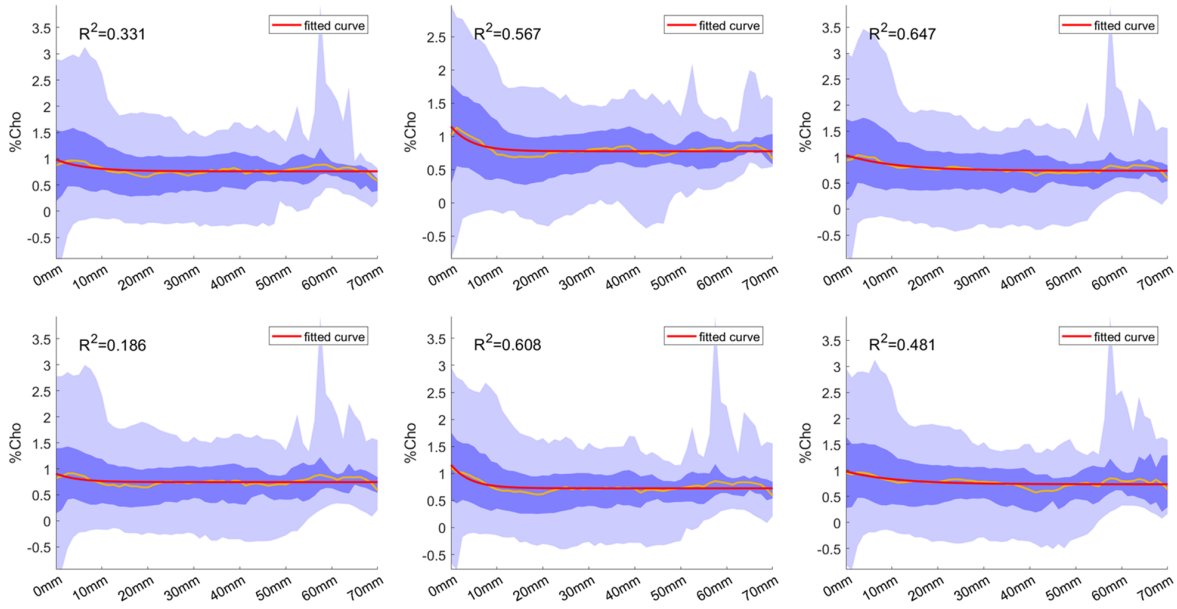

Figure 11.12: MRS-FSD fitting results for %Cho. Each plot corresponds to one fold of the 6-fold CV scheme used, where at each time one-sixth of the examinations were excluded from the data used to fit the model. Each plot shows the percentiles per distance to the STV, similarly to what is described in the legend of Figures 11.10 and 11.11. The fitted curves, shown in red, were fitted to the median curves, shown in yellow. The fitting was performed after feature rescaling as described in Section 11.3.6 (after rescaling 0 corresponds to the 5th percentile of the feature in healthy volunteers, and 1 to the 95th percentile). Each plot shows the corresponding coefficient of determination (R^2).

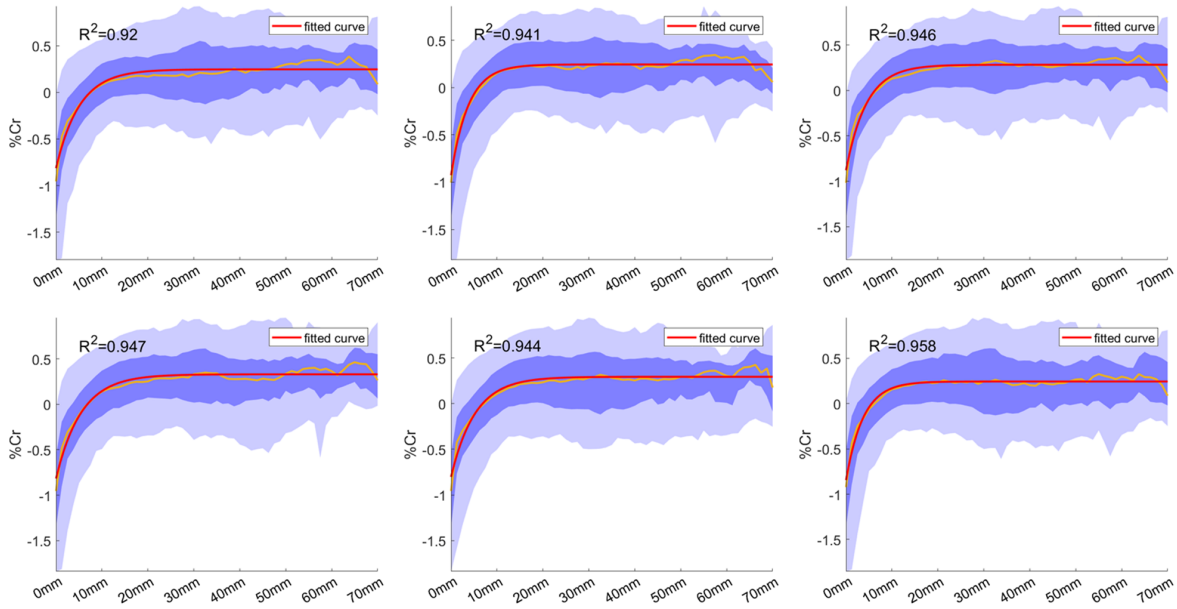

Figure 11.13: MRS-FSD fitting results for %Cr. Each plot corresponds to one fold of the 6-fold CV scheme used, where at each time one-sixth of the examinations were excluded from the data used to fit the model. Each plot shows the percentiles per distance to the STV, similarly to what is described in the legend of Figures 11.10 and 11.11. The fitted curves, shown in red, were fitted to the median curves, shown in yellow. The fitting was performed after feature rescaling as described in Section 11.3.6 (after rescaling 0 corresponds to the 5th percentile of the feature in healthy volunteers, and 1 to the 95th percentile). Each plot shows the corresponding coefficient of determination (R^2).

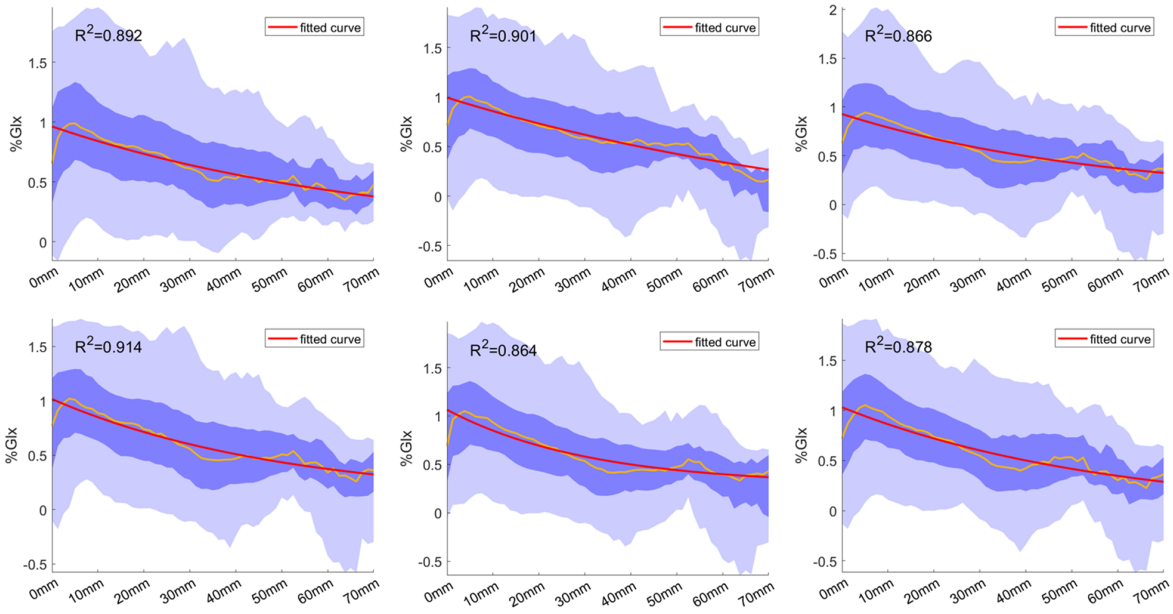

Figure 11.14: MRS-FSD fitting results for %Glx. Each plot corresponds to one fold of the 6-fold CV scheme used, where at each time one-sixth of the examinations were excluded from the data used to fit the model. Each plot shows the percentiles per distance to the STV, similarly to what is described in the legend of Figures 11.10 and 11.11. The fitted curves, shown in red, were fitted to the median curves, shown in yellow. The fitting was performed after feature rescaling as described in Section 11.3.6 (after rescaling 0 corresponds to the 5th percentile of the feature in healthy volunteers, and 1 to the 95th percentile). Each plot shows the corresponding coefficient of determination (R^2).

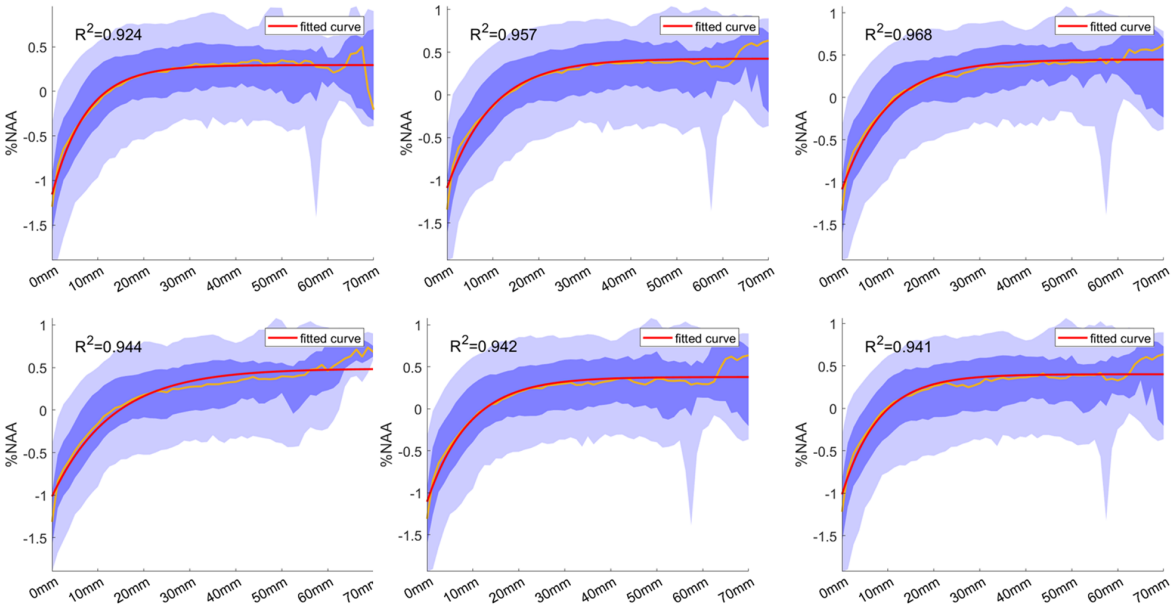

Figure 11.15: MRS-FSD fitting results for %NAA. Each plot corresponds to one fold of the 6-fold CV scheme used, where at each time one-sixth of the examinations were excluded from the data used to fit the model. Each plot shows the percentiles per distance to the STV, similarly to what is described in the legend of Figures 11.10 and 11.11. The fitted curves, shown in red, were fitted to the median curves, shown in yellow. The fitting was performed after feature rescaling as described in Section 11.3.6 (after rescaling 0 corresponds to the 5th percentile of the feature in healthy volunteers, and 1 to the 95th percentile). Each plot shows the corresponding coefficient of determination (R^2).

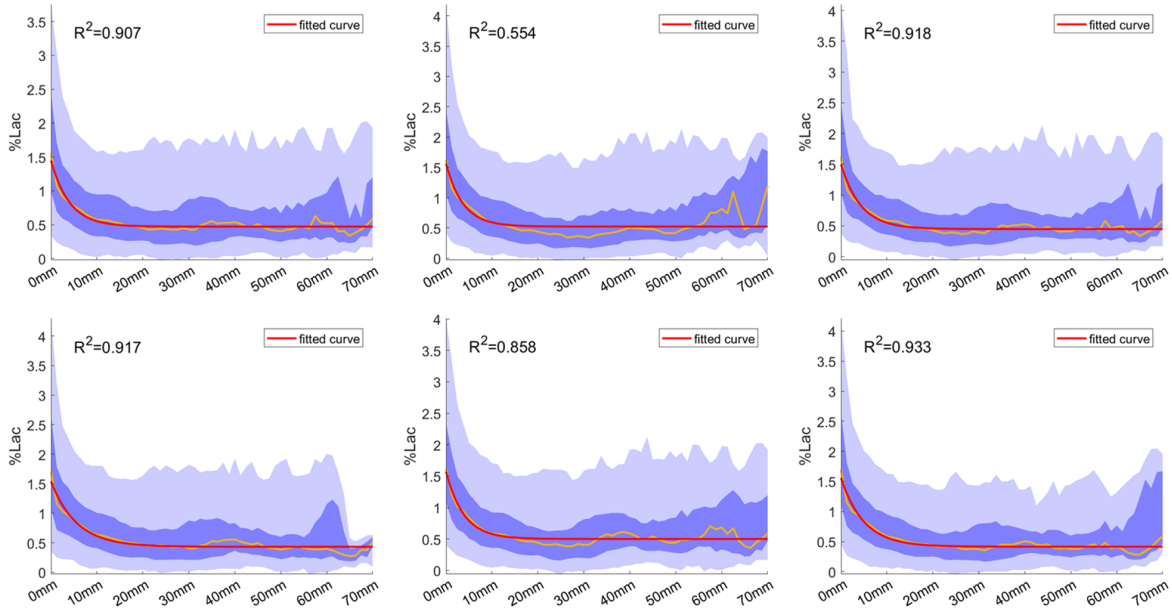

Figure 11.16: MRS-FSD fitting results for %Lac. Each plot corresponds to one fold of the 6-fold CV scheme used, where at each time one-sixth of the examinations were excluded from the data used to fit the model. Each plot shows the percentiles per distance to the STV, similarly to what is described in the legend of Figures 11.10 and 11.11. The fitted curves, shown in red, were fitted to the median curves, shown in yellow. The fitting was performed after feature rescaling as described in Section 11.3.6 (after rescaling 0 corresponds to the 5th percentile of the feature in healthy volunteers, and 1 to the 95th percentile). Each plot shows the corresponding coefficient of determination (R^2).

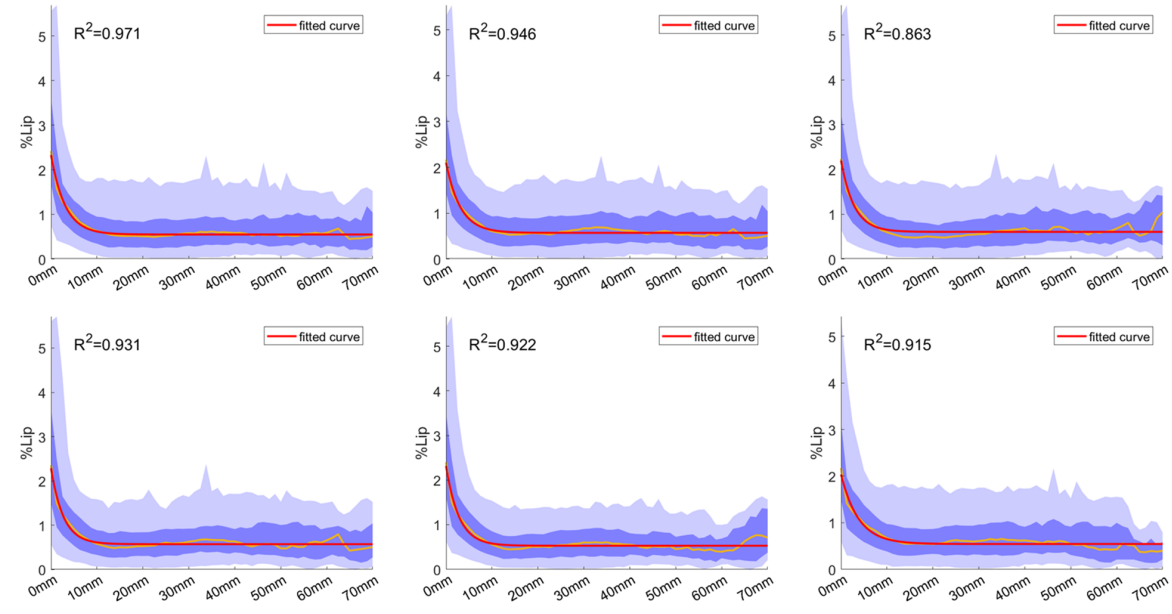

Figure 11.17: MRS-FSD fitting results for %Lip. Each plot corresponds to one fold of the 6-fold CV scheme used, where at each time one-sixth of the examinations were excluded from the data used to fit the model. Each plot shows the percentiles per distance to the STV, similarly to what is described in the legend of Figures 11.10 and 11.11. The fitted curves, shown in red, were fitted to the median curves, shown in yellow. The fitting was performed after feature rescaling as described in Section 11.3.6 (after rescaling 0 corresponds to the 5th percentile of the feature in healthy volunteers, and 1 to the 95th percentile). Each plot shows the corresponding coefficient of determination (R^2).

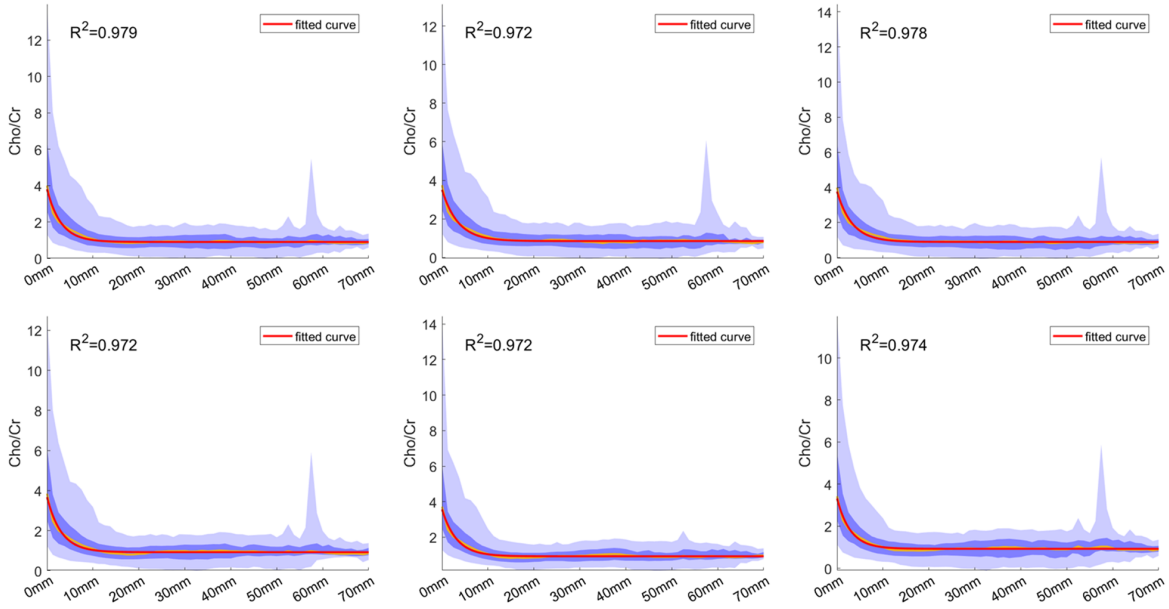

Figure 11.18: MRS-FSD fitting results for Cho/Cr. Each plot corresponds to one fold of the 6-fold CV scheme used, where at each time one-sixth of the examinations were excluded from the data used to fit the model. Each plot shows the percentiles per distance to the STV, similarly to what is described in the legend of Figures 11.10 and 11.11. The fitted curves, shown in red, were fitted to the median curves, shown in yellow. The fitting was performed after feature rescaling as described in Section 11.3.6 (after rescaling 0 corresponds to the 5th percentile of the feature in healthy volunteers, and 1 to the 95th percentile). Each plot shows the corresponding coefficient of determination (R^2).

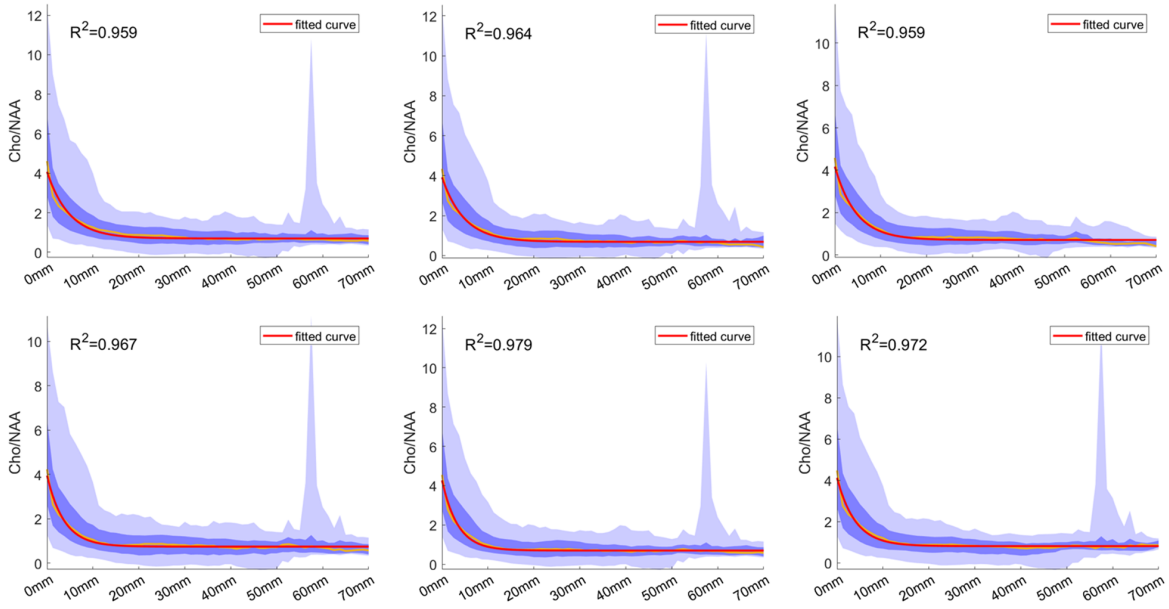

Figure 11.19: MRS-FSD fitting results for Cho/NAA. Each plot corresponds to one fold of the 6-fold CV scheme used, where at each time one-sixth of the examinations were excluded from the data used to fit the model. Each plot shows the percentiles per distance to the STV, similarly to what is described in the legend of Figures 11.10 and 11.11. The fitted curves, shown in red, were fitted to the median curves, shown in yellow. The fitting was performed after feature rescaling as described in Section 11.3.6 (after rescaling 0 corresponds to the 5th percentile of the feature in healthy volunteers, and 1 to the 95th percentile). Each plot shows the corresponding coefficient of determination (R^2).

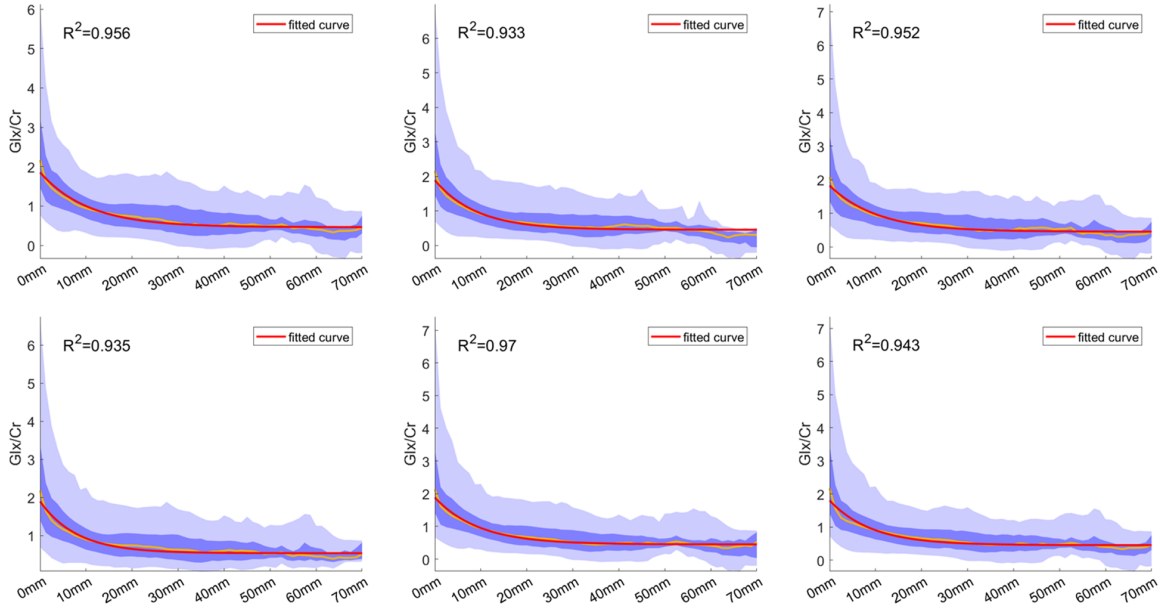

Figure 11.20: MRS-FSD fitting results for Glx/Cr. Each plot corresponds to one fold of the 6-fold CV scheme used, where at each time one-sixth of the examinations were excluded from the data used to fit the model. Each plot shows the percentiles per distance to the STV, similarly to what is described in the legend of Figures 11.10 and 11.11. The fitted curves, shown in red, were fitted to the median curves, shown in yellow. The fitting was performed after feature rescaling as described in Section 11.3.6 (after rescaling 0 corresponds to the 5th percentile of the feature in healthy volunteers, and 1 to the 95th percentile). Each plot shows the corresponding coefficient of determination (R^2).

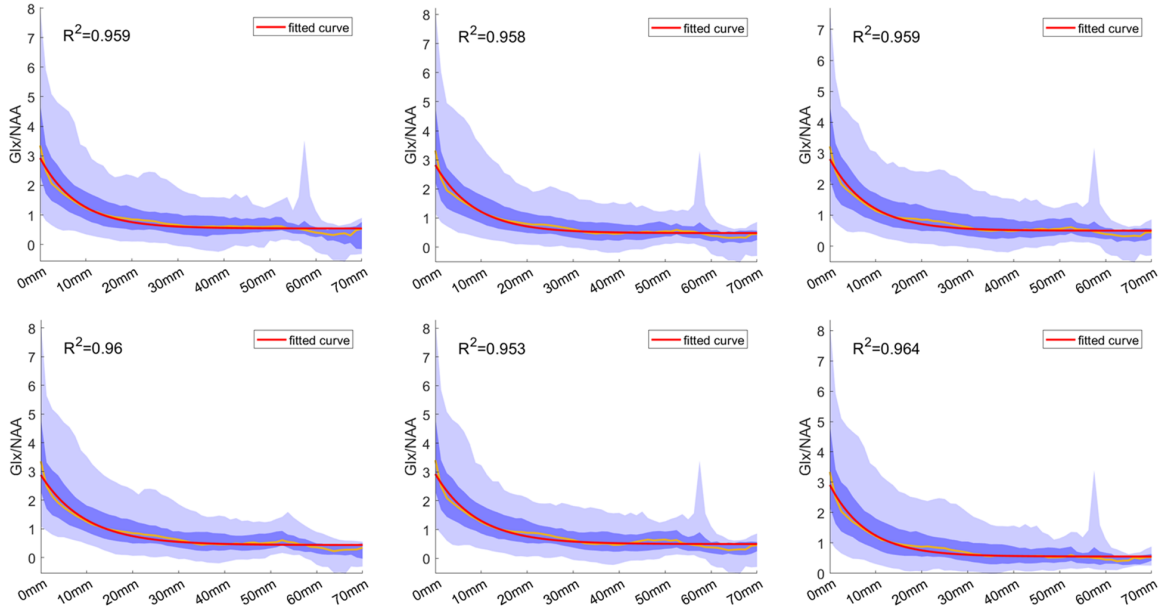

Figure 11.21: MRS-FSD fitting results for Glx/NAA. Each plot corresponds to one fold of the 6-fold CV scheme used, where at each time one-sixth of the examinations were excluded from the data used to fit the model. Each plot shows the percentiles per distance to the STV, similarly to what is described in the legend of Figures 11.10 and 11.11. The fitted curves, shown in red, were fitted to the median curves, shown in yellow. The fitting was performed after feature rescaling as described in Section 11.3.6 (after rescaling 0 corresponds to the 5th percentile of the feature in healthy volunteers, and 1 to the 95th percentile). Each plot shows the corresponding coefficient of determination (R^2).

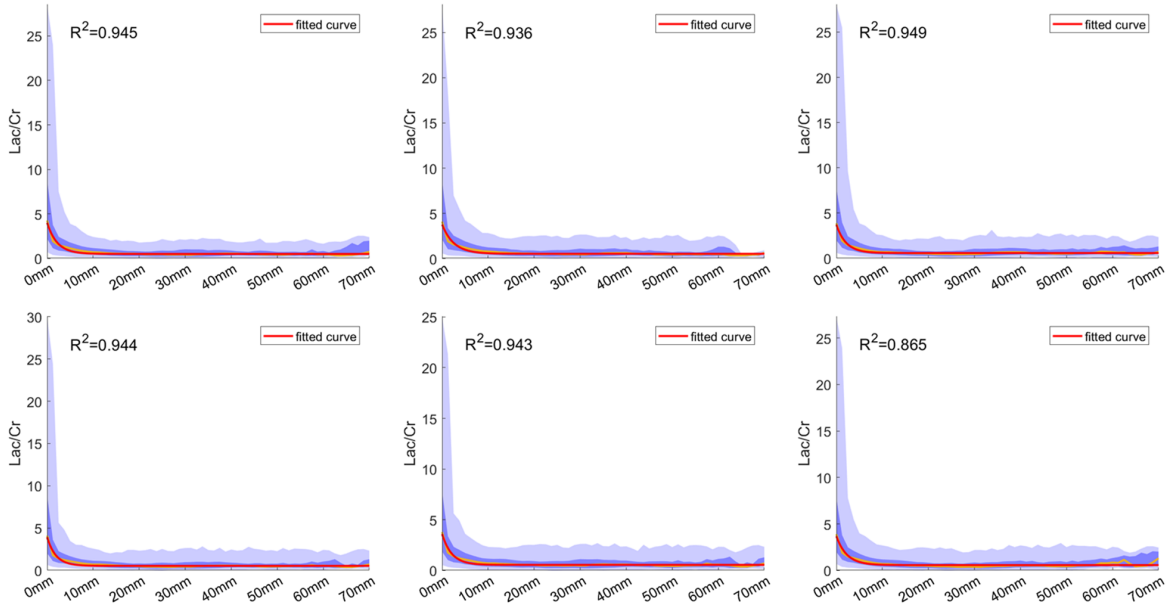

Figure 11.22: MRS-FSD fitting results for Lac/Cr. Each plot corresponds to one fold of the 6-fold CV scheme used, where at each time one-sixth of the examinations were excluded from the data used to fit the model. Each plot shows the percentiles per distance to the STV, similarly to what is described in the legend of Figures 11.10 and 11.11. The fitted curves, shown in red, were fitted to the median curves, shown in yellow. The fitting was performed after feature rescaling as described in Section 11.3.6 (after rescaling 0 corresponds to the 5th percentile of the feature in healthy volunteers, and 1 to the 95th percentile). Each plot shows the corresponding coefficient of determination (R^2).

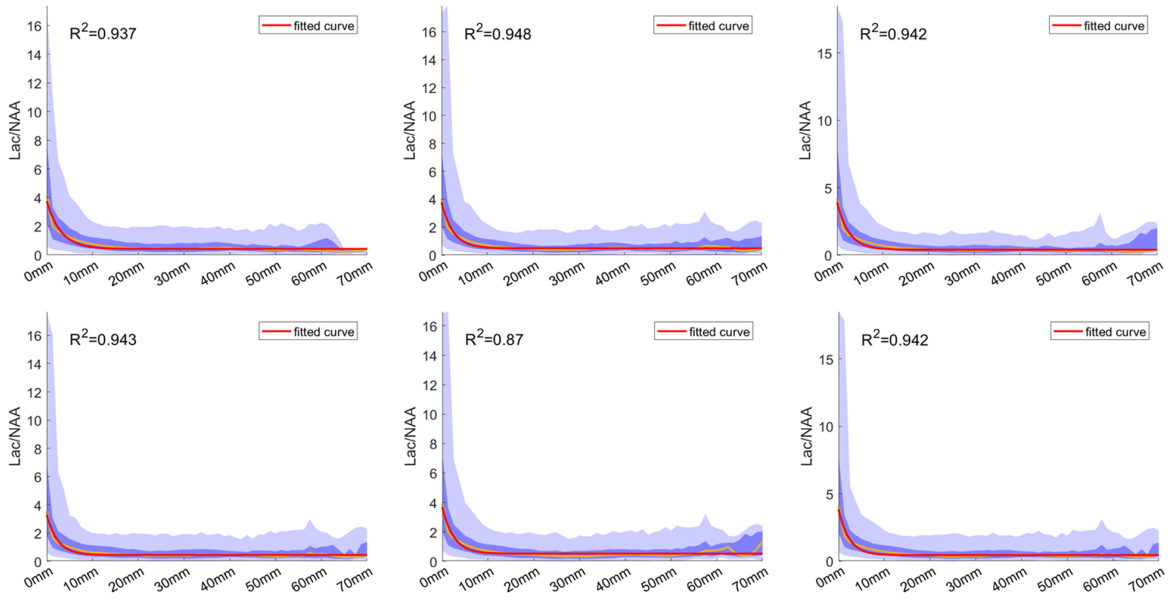

Figure 11.23: MRS-FSD fitting results for Lac/NAA. Each plot corresponds to one fold of the 6-fold CV scheme used, where at each time one-sixth of the examinations were excluded from the data used to fit the model. Each plot shows the percentiles per distance to the STV, similarly to what is described in the legend of Figures 11.10 and 11.11. The fitted curves, shown in red, were fitted to the median curves, shown in yellow. The fitting was performed after feature rescaling as described in Section 11.3.6 (after rescaling 0 corresponds to the 5th percentile of the feature in healthy volunteers, and 1 to the 95th percentile). Each plot shows the corresponding coefficient of determination (R^2).

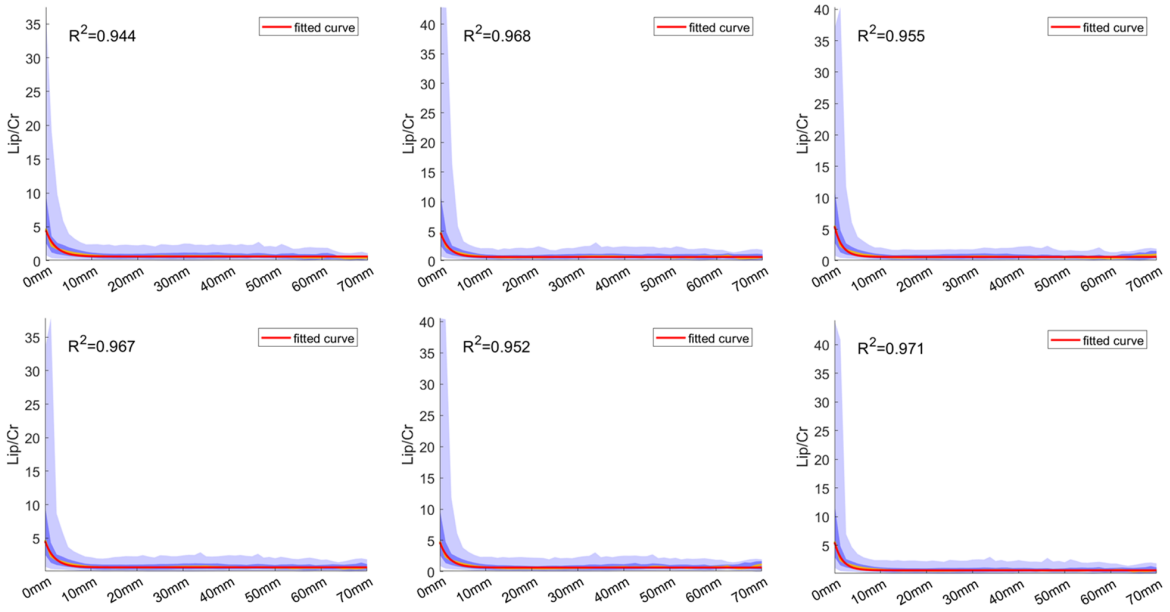

Figure 11.24: MRS-FSD fitting results for Lip/Cr. Each plot corresponds to one fold of the 6-fold CV scheme used, where at each time one-sixth of the examinations were excluded from the data used to fit the model. Each plot shows the percentiles per distance to the STV, similarly to what is described in the legend of Figures 11.10 and 11.11. The fitted curves, shown in red, were fitted to the median curves, shown in yellow. The fitting was performed after feature rescaling as described in Section 11.3.6 (after rescaling 0 corresponds to the 5th percentile of the feature in healthy volunteers, and 1 to the 95th percentile). Each plot shows the corresponding coefficient of determination (R^2).

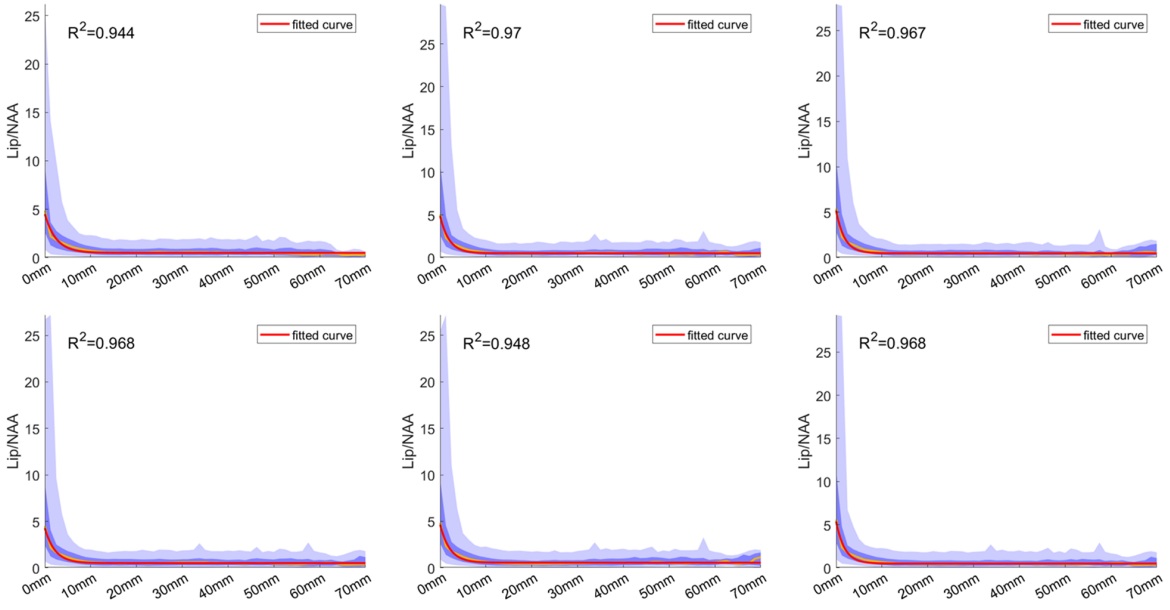

Figure 11.25: MRS-FSD fitting results for Lip/NAA. Each plot corresponds to one fold of the 6-fold CV scheme used, where at each time one-sixth of the examinations were excluded from the data used to fit the model. Each plot shows the percentiles per distance to the STV, similarly to what is described in the legend of Figures 11.10 and 11.11. The fitted curves, shown in red, were fitted to the median curves, shown in yellow. The fitting was performed after feature rescaling as described in Section 11.3.6 (after rescaling 0 corresponds to the 5th percentile of the feature in healthy volunteers, and 1 to the 95th percentile). Each plot shows the corresponding coefficient of determination (R^2).

12 | Overall Discussion & Outlook

12.1 Discussion

This thesis addressed several aspects that currently limit the use of proton MRS in clinical practice, namely the need for extensive user expertise and the considerable amount of time required to process and analyze MRS data. Before presenting the research results of this doctoral thesis, the first part of thesis introduced the reader to the main basic concepts underlying Nuclear Magnetic Resonance (NMR), Magnetic Resonance Imaging (MRI) and Spectroscopy (MRS) (chapter 2), MRS data processing & machine learning (chapter 3) and the clinical use of proton MRS for the characterization of brain tumors (chapter 4). The research results were presented in the second and third parts of this thesis.

12.1.1 Part II - quality management of MRS and automation of quality control

The second part of the thesis starts with chapter 5, where several of the aspects that need to be considered to guarantee reproducible and unambiguous results in *in vivo* proton MRS were discussed. More specifically, the chapter covered different important details on the use of phantom experiments for quality assurance, described different methods for detecting errors in the acquisition and quantification, and presented different strategies that can be used to improve MRS data quality. Proper quality management is essential for the simplification of MRS, given that signal artifacts and bad quality data are some of the main factors that limit the clinical use of MRS and require (neuro)radiologists/clinicians to have solid expertise in MRS. In other words, if quality assurance is performed correctly by the technical staff, and the software used to handle MRS data takes care of controlling the quality of the data, not only interpretation errors can be prevented but clinicians may focus on the purely clinical aspects of MRS.

Continuing the topic of quality management, chapter 6 presents a method for the automatic quality control of MRS data. The method corresponds to an automatic classifier (Random Forests [Breiman 2001]) that was trained to mimic the subjective judgment of an expert regarding the quality of MR spectra. The results showed that the developed approach is able to classify several hundreds of spectra almost instantaneously and with the same accuracy as an expert. Given that the method outputs an estimation of the probability of a given spectrum being "accepted" by a rater, i.e. the output is continuous, it is possible to perform quality filtering using different thresholds. This feature can be useful for specific situations where, for instance, the quality requirements are particularly high, or alternatively, in cases where low-quality spectra are still able to provide an answer for the clinical question at hand.

Another important aspect of automatic quality control is the consistency and objectiveness of the automatic approach. As the results showed, different raters often disagree in their assessments and even the same rater is often inconsistent when repetitively rating the quality

of the spectra. This shows that, beyond the time-saving potential and the facilitation of the analysis that results from removing "bad" quality data, automatic quality control contributes to a more consistent and objective assessment of the quality of MR spectra. Moreover, its performance can always be improved using new learning approaches and classification methods, incorporating more features, etc.. The subjectivity of the manual assessment of the signal quality in MRS, and the dependency of the classification performance on the type of features and classification method used, are well covered in a recently published paper by Kyathanahally et al. [Kyathanahally 2017].

Following the development of the presented automatic quality control method, a new approach using the concept of active learning was presented that aimed at improving the labeling efficiency of MRSI data (chapter 7). The study showed that it is possible to drastically reduce the number of spectra that need to be labeled without sacrificing the performance of the classifier in distinguishing "good" from "bad" quality data. Moreover, it allowed to better understand the relationship between classification performance and the number of training examples, as well as to conclude that the spectra that contain the most information for training automatic classifiers are the ones whose classification is more uncertain. This is due to the fact that these spectra are the ones that ultimately will define the hyperplanes in the feature space corresponding to the class separation boundaries. Finally, the results demonstrated that the most efficient strategy to label MR spectra corresponds to a combination of diversity and uncertainty, i.e., the classifier learns more efficiently if the most uncertain spectra from as many different examinations as possible are selected.

12.1.2 Part III - assisting the interpretation of brain tumor MRSI data

After the initial part dedicated to quality aspects of MRS, the doctoral work changed the focus to methods that aim at assisting in the interpretation of MR spectra. The methods were developed for the specific case of brain tumors, however, the concepts used are general and may be applied to other organs and, possibly, other (focal) diseases. The different methods have all in common that they use information from previously analyzed data to help to interpret new spectra. Moreover, the different methods transform each new MRSI acquisition into an easily interpretable map that places the new examination in the context of the already known data. Consequently, not only these methods lower the requirements in terms of expertise required to properly interpret the results, they can also be very interesting for experienced users, allowing them to easily incorporate more information into the analysis of MRS examinations. Regardless of the experience, given that proper interpretation of MRS examinations can be extremely time-consuming, the assistance in interpreting MRS data is expected to have a considerable time-saving potential.

Going now into more detail on the different methods presented, chapter 8 showed the results of using convex NMF to the analysis of data from GBM and metastasis patients. The results showed that several interesting aspects of the data can be highlighted using 3 sources to describe the MRSI data of these two types of brain tumors. The cNMF with 3 sources converges to a solution where one source represents normal brain spectra, the second source represents solid tumor/necrotic tissue, and a third source was interpreted as a tumor infiltration pattern (high Cho/NAA but only slightly increased Cho/Cr, presence of lactate and no mobile lipids). The results of the study seem to indicate that the infiltration source is usually more prevalent in GBM patients, however the p-value of the difference in the mean mixing matrix for the infiltration source between GBM and metastasis is only 0.058, meaning that more data would

be needed to confirm or reject the null hypothesis. Despite this, the standard deviation of the mixing matrix associated with the infiltration source was significantly higher in GBM patients. The results of this study confirmed the findings reported by other researchers, describing that the main differences between brain metastasis and GBM are found in the tumor periphery ([Server 2010, Wijnen 2012]), given that the sources that showed significant differences between the two types of tumors are usually found in the periphery of these tumors.

The experience gained with the project of chapter 8 showed that the use of reference spectra can assist in the interpretation of MRS data but that different aspects of such approaches need to be tackled. The first one is related to a known effect of MRS acquisitions: the chemical shift displacement artifact (CSDA). Due to the differences in the resonance frequency of the main resonances of Cho and NAA, the selection profiles of these two metabolites are shifted relative to each other. Consequently, even in a homogeneous phantom, the Cho/NAA ratio is not constant in the different voxels of the MRSI grid (This artifact is discussed in more detail in 5.5.4). Given that increased Cho/NAA is a marker for tumor infiltration, and a distinguishing feature of the infiltration source of chapter 8, the presented results showed a tendency to identify red-source in the upper (frontal) voxels of the grid, where due to the CSDA, Cho/NAA was "artificially" increased. This problem led to the development of the method presented in section 5.5.8, which consists in the use of phantom measurements with the same acquisition parameters of the acquired data to correct the selection profile and CSDA. The method was included in the processing pipeline of the methods developed posteriorly to prevent the incorrect identification of infiltration features (see, for instance, Figure 10.2 of chapter 10).

Another important aspect that was missing in the processing pipeline used to produce the results of chapter 8 was quality filtering. In this case, the problem had already been identified, however, the method presented in chapter 6 was still being developed at that time and consequently, it was decided not to include it at that moment.

Finally, one of the main problems identified with cNMF was the difficulty in defining objective criteria to decide the number of sources to use. This motivated the interest in X-means ([Pelleg 2000]), a method that despite being substantially different from NMF, can be used in a similar way since it is able to identify groups of spectra that share similar spectroscopic properties. More importantly, in X-means the identification of the number of clusters is implicit to the method and performed based on a hierarchical approach, where at every the time the algorithm decides based on the Bayesian Information Criterion (BIC) whether to divide a group into two subgroups or not. After identifying the different groups, the average spectra (prototype spectra) of each cluster can be used to assign new spectra to each group. The use of the prototype spectra in X-means to interpret new spectra, despite different, can be compared to the use of the different sources retrieved using cNMF for interpreting new cases. In both cases, the reference spectra used correspond to an average over a certain group of spectra, however, cNMF is a factorization method and the selected sources correspond to basis vectors that may be used to reconstruct new cases. This means that fuzzy clustering is implicit to cNMF and that cNMF is particularly suitable to handle low-resolution data where the majority of voxels contain mixtures of tissues. In the case of X-means, fuzzy clustering can be achieved using the value of a given distance metric applied to evaluate the similarity between a new spectrum and the reference spectra. However, in this thesis, the prototype spectra identified with X-means were simply used to identify the "closest" cluster and not explored for fuzzy-clustering.

The first results obtained with X-means, which included selection profile correction and automatic quality control, were presented in chapter 9. The MRSI segmentation achieved with this method revealed a total of 9 different groups and allowed to identify more pronounced

features than the previous method, such as spectroscopy patterns with very high Cho/Cr but almost any mobile lipids. Moreover, in this case MRI segmentation was available and could be used to compare the segmentation results obtained with MRI and MRSI. In these preliminary results, it could be seen that despite a rough agreement between imaging and spectroscopy, both modalities carry different information. Despite the interesting results, a new problem arose with the use of X-Means: depending on the dataset, features, processing pipeline, etc. the number of clusters identified by X-means changed considerably. This meant that the proposed approach could not be used clinically since the number and type of clusters would change considerably depending on the dataset used for clustering to define the reference spectra. This problem led to the development of BraTSIS (Brain Tumor Spectroscopic Imaging Segmentation).

BraTSIS (chapter 10) was developed with the aim of defining clear rules that could standardize the identification of the main types of spectra seen in MRSI data of GBM patients. Therefore, the use of well-defined and widely available references to define the different groups was one of the main concerns since the method had to be reproducible between different centers. Consequently, MRI segmentation and data from healthy controls were the two selected references. To guarantee the transparency of the method and inspired by the majority of guidelines used in clinical practice, a decision tree was defined that could be easily applied in any center and improved over different iterations. This ability to improve it iteratively, and possibly through the collaboration between different centers, was seen as an essential aspect to guarantee that BraTSIS adequately addresses the needs of clinical users. The structure of the tree was defined considering the results obtained with X-Means. Compared with the method of the previous chapter, an extended feature set was used that included, besides other metabolite ratios, Glx/NAA and Glx/Cr. This allowed identifying a novel type of metabolic pattern (named High Glx), that besides being widely present in the analyzed dataset hasn't been described in previous literature. The High Glx pattern is characterized by a close to normal brain spectrum, with the exception of extremely elevated Glx levels, which may be related to inflammatory processes associated with the tumor. The results showed that BraTSIS can successfully identify the main types of spectra consistently and may be a starting point for the standardization of the analysis of MRSI data in GBM patients.

Finally, the work presented in chapter 11 originated from several questions that arose during the development of BraTSIS. BraTSIS explores the concept that in different types of patterns, different features represent the differentiating factor needed to identify them. This is the case, for instance, with spectra found in tissue necrosis, which can be easily identified by the very high mobile lipid content and very low levels of creatine and NAA. However, despite this assumption, it was not clear the behavior of different metabolite ratios in the transition between different groups. While trying to answer this and other questions, it was noticed that, relatively to the tumor border, different tissues are usually found at different distances of the border. More precisely, healthy brain tissue is found usually in regions very far from the tumor, infiltrated tissue in regions outside the tumor border but very close to the tumor, solid/active tumor in the peripheral zone of the tumor, and necrosis in the core of the solid tumor. This general relation between tissue types and the distance to solid tumor – which may certainly find numerous exceptions – was the motivation for analyzing the relationship between the distance to the tumor border visible in structural MRI and several metabolite ratios.

The analysis of the relationship between the distance to tumor and the different metabolite ratios was the starting point for the concept of the Expected Distance to Tumor (EDT) maps, presented in the same chapter (11). After identifying a clear relationship between almost every metabolite ratio and the distance to the tumor border, it was hypothesized that a model trained

to learn the relationship between metabolic information and the distance to solid tumor could identify tumor infiltration as regions with a short expected distance to solid tumor, compared with regions with normal spectroscopic patterns, which would be expected to be found at large distances to solid tumor. Consequently, a regression model (Regression Forests [Liaw 2002]) was trained to estimate the distance to solid tumor based on spectroscopic features of a single voxel. The results showed that the different metabolic patterns are associated to different EDT values. More importantly, the ordering of the spectra by the predicted distances, i.e. the EDT value, corresponds to an ordering of different metabolic patterns ranging from normal brain spectra to necrosis, with a clear association between malignant features, such as high Cho/NAA and high Cho/Cr, and short EDT values. This method that was developed at the end of the Ph.D. programme is seen – in the perspective of the author – as a very interesting solution to assist the interpretation of brain tumor MRS data. Not only can it be easily reproduced by different centers since the definition of tumor borders in which it relies is one of the most well-defined tasks in neuroradiology, it does not require the definition of a fixed number clusters (as in X-means[1] or cNMF). Moreover, it provides a linear scale that orders the majority of metabolic patterns seen in GBM patients in terms of their usual distance to the tumor borders, which seems to correspond to the ordering of metabolic patterns in terms of their malignancy. The output is also of easy interpretation: an EDT value of 10 mm, for instance, can be easily interpreted as a metabolic pattern that is usually found at a distance of 10 mm from the tumor border. Finally, the identification of a border for tumor-related abnormalities may be performed simply using as reference the EDT values found in regions known to contain only healthy brain tissue.

12.2 Overview and outlook

This thesis was part of a European Marie-Curie project named TRANSACT (TRANSforming magnetic resonance spectroscopy into A Clinical Tool). In the context of this project, the doctoral work here presented contributed to solving some of the strongest obstacles for the clinical use of MRS. The different results presented here confirmed that it is possible to greatly simplify the processing and analysis of MRS using machine learning. Moreover, in SpectrIm it was possible to define fully automatic processing pipelines that allow clinical users to focus solely on the interpretation of the final results.

One of the main topics of this thesis was the automation of the quality control of MRS data. The results showed that it is not only possible to develop classifiers as accurate as expert users, they can classify huge amounts of data almost instantaneously and consistently. Moreover, a new optimized approach to label the training MR spectra was also presented. To continue the work of the development of automatic quality control classifiers, it could be interesting to compare the performance of such methods for the identification of different types of artifacts and identify in which cases it could be further improved. It is expected, for instance, that less common artifacts may be recognized with a poorer performance than more common ones. Such a study would not only allow to improve the overall classification performance but also to better understand the limitations of the method. Another important aspect that was not analyzed in this thesis was the relationship between spectral quality and the quantification error. The

[1] In the case of X-means, the determination of the number of clusters is done automatically, meaning that, in theory, the definition of the number of clusters is not a problem for the user. However, in practice, different solutions may be obtained as a result of minor changes in the dataset and a considerable amount of work may be required to obtain meaningful and reproducible results.

signal quality control method presented is based solely on the human perception of the quality of spectra and aims at identifying artifacts that not only lead to quantification errors, but also to wrong estimations of the quantification error. Given that the results' uncertainty induced by signal artifacts should be the main factor driving the decision on whether to reject or accept a given spectrum, it could be interesting to develop machine learning based systems that could robustly and accurately predict the quantification error in the presence of signal artifacts. In a further step, the error-prediction could move to an estimation of the impact that the quality of the data would have on the uncertainty of the clinical interpretation of the data.

On the topic of selection profile correction, it was showed that it can be easily done using simple phantom acquisitions. However, it should be possible to more precisely calculate the correction factors simply using the information from the pulse sequence, namely the pulse shapes and bandwidths. This is an artifact that, besides having a simple solution, is currently affecting examinations used in clinical practice and can lead to wrong interpretation of MRS data. Therefore, it would be important to develop and distribute correction approaches for this effect that could be brought into clinical practice in the near future.

Moving now to the third part of the thesis, different strategies were explored to assist the analysis of brain tumor MRSI data: cNMF, X-Means, BraTSIS and EDT maps. In all cases, the results presented in this thesis showed that automatic quality control and selection profile/CSDA correction are essential for the success of these approaches. The presented approaches showed that, using such methods it is possible to highlight interesting aspects of the MRSI data of brain tumors. It is now necessary to validate the different results obtained with the developed methods using follow-up data and histology. Their validation should also enable to further improve and adapt them to maximize their clinical utility. Moreover, it would be extremely interesting to evaluate some of these methods using high spatial resolution MRSI data, acquired with sequences such as EPSI [Mulkern 2001] or SPICE [Lam 2015, Ma 2016].

For several years, the clinical value of MRS has been shown in different research papers, some of which were covered in this thesis, namely in chapter 4. However, the complexity of MRS has greatly limited the use of this technique causing it to be only available in a few specialized centers. The limited amount of users means less motivation for the main vendors to develop new and more attractive systems that improve and simplify the handling of MRS data. Moreover, the existence of less clinical users means also that the knowledge and understanding of the associations between the information provided by MRS and distinct diseases and conditions evolves slowly. Overall, the combination of all these interconnected effects has limited the clinical application and applicability of MRS greatly.

Despite this, the recent technological developments in MRI scanner hardware and MRS pulse sequences lead to an improvement in the quality of MRS data. More importantly, the democratization of machine learning and data sciences, and the increasing availability of better and more advanced IT solutions means that it is now easier than ever to handle complex data. The ability to handle complex data in an easy fashion and the growing interest in information-rich data are expected to contribute to a shift in the perception of MRS, from a situation where its complexity is seen as limitation for the clinical use, to another where the complexity is regarded as an advantage that enables clinicians to perform tasks not possible with other techniques. Consequently, the clinical use of MRS is expected to rise in the next years, accompanying, at least, the annual growth in MRI examinations per capita (see Figure 12.1).

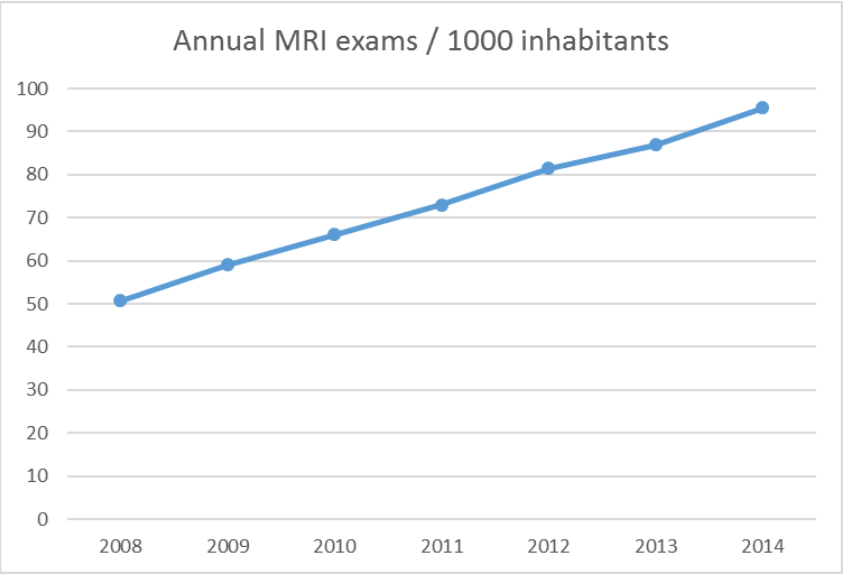

Figure 12.1: Evolution of the number of MRI examinations per year, in Europe. Data source: OECD (2017), Magnetic resonance imaging (MRI) exams (indicator). doi: 10.1787/1d89353f-en.

A | Supervised Master Thesis

A.1 Segmentation-Guided MRS Analysis of Brain Tumor Patients

- Student: Martin Pletscher
- Supervisors: Nuno Barros, Raphael Meier
- First & second examiner: Mauricio Reyes, Johannes Slotboom

Magnetic resonance imaging (MRI) is considered the method of choice for glioblastoma multiforme diagnosis and follow-up. MRI provides structural information and plays a crucial role in progression monitoring, treatment planning, and analysis of brain tumors. Two-dimensional measurement of tumor diameters is considered the gold standard for neuro-oncologic therapy response assessment as of today. In recent years magnetic resonance spectroscopy (MRS) is used increasingly to gain additional information on a metabolic level to improve tumor diagnosis. The additional information comes at the cost of a considerable amount of experience needed by clinicians to translate spectral information into histopathological tissue information. Therefore, the aim of this work was to investigate a tissue segmentation method to help facilitate the use of MRS in neuro-oncologic glioblastoma assessment by automating the translation process. The 22 pre-operative data sets from patients with histologically confirmed glioblastoma multiforme contained four MRI sequences (T_1, T_{1c}, T_2 and FLAIR) and ^{1}H MRSI scan data (2D-PRESS, CHESS, TE=135 ms, TR=1500 ms) acquired at a native resolution of 12×12 and interpolated to 32×32 voxels. The data sets were acquired on 1.5T Siemens scanners (Area, Avanto). The measurements were performed conforming to local and national ethical regulations: all patients gave their informed consent to use the data for scientific purposes. In the first step of the proposed method, 17 randomly selected data sets were pre-processed with jMRUI's Spectrim plugin (residual water removal HLSVD, qualityfiltering, frequency-shift correction, and autophasing). After QUEST quantification a feature set of metabolite-ratios Cho/Cr, NAA/Cho, NAA/Cr, (Lip1.3+Lac)/Cr, (Lip1.3+Lac)/NAA, (Lip1.3+Lac)/Cho was extracted from the 17 data sets. The feature set was then clustered by the x-means algorithm. The clustering revealed nine characteristic sources (Necrosis 1, Necrosis 2, Necrosis + Active Tumor, Active Tumor, Active Tumor + Infiltration, Infiltration, Normal WM1, Normal WM2, and Normal GM) which were labeled and ordered by an experienced spectroscopist. In the second step, the nine sources were used to segment the remaining five test data sets which were excluded from the clustering. The result of the segmentation was visualized as a map with colors ranging from red (malignant) to blue (healthy) for each type of tissue found. Additionally, a visual and statistical comparison to a structural segmentation of the same data sets, generated by the Brain Tumor Image Analysis (BraTumIA) software, was performed. The resulting segmentation maps show a

good visual resemblance of the structural segmentation. A quantitative analysis with the Dice similarity coefficient yielded a mean spatial overlap of 0.62 ± 0.29 for the malignant regions and 0.71 ± 0.20 for healthy regions, both areas excluding the Infiltration and Edema classes. Only low correspondence of 0.22 ± 0.14 was found between Infiltration and Edema. We conclude that the proposed method has the potential to enhance the clinical use of MRS in neuro-oncologic glioblastoma assessment. It facilitates and largely automatizes the translation of spectral information into histopathological tissue information. In future research, the method could be applied to other types of brain tumors to gain a better understanding of the differences between those.

Bibliography

[Aggarwal 2001] Charu C Aggarwal, Alexander Hinneburg and Daniel A Keim. *On the surprising behavior of distance metrics in high dimensional spaces*. In ICDT, volume 1, pages 420–434. Springer, 2001. (Cited on page 151.)

[Albrecht 2010] Jan Albrecht, Marta Sidoryk-Węgrzynowicz, Magdalena Zielińska and Michael Aschner. *Roles of glutamine in neurotransmission*. Neuron Glia Biology, vol. 6, no. 04, pages 263–276, 2010. (Cited on page 55.)

[Aldave 2013] Guillermo Aldave, Sonia Tejada, Eva Pay, Miguel Marigil, Bartolomé Bejarano, Miguel A. Idoate and Ricardo Díez-Valle. *Prognostic value of residual fluorescent tissue in glioblastoma patients after gross total resection in 5-aminolevulinic acid-guided surgery*. Neurosurgery, vol. 72, no. 6, pages 915–920, 2013. (Cited on page 172.)

[Andronesi 2012] O. C. Andronesi, G. S. Kim, E. Gerstner, T. Batchelor, a. a. Tzika, V. R. Fantin, M. G. Vander Heiden and a. G. Sorensen. *Detection of 2-Hydroxyglutarate in IDH-Mutated Glioma Patients by In Vivo Spectral-Editing and 2D Correlation Magnetic Resonance Spectroscopy*. Science Translational Medicine, vol. 4, no. 116, pages 116ra4–116ra4, 2012. (Cited on page 57.)

[Andronesi 2013] Ovidiu C. Andronesi, Otto Rapalino, Elizabeth Gerstner, Andrew Chi, Tracy T. Batchelor, Dan P. Cahill, a. Gregory Sorensen and Bruce R. Rosen. *Detection of oncogenic IDH1 mutations using magnetic resonance spectroscopy of 2-hydroxyglutarate*. Journal of Clinical Investigation, vol. 123, no. 9, pages 3659–3663, 2013. (Cited on page 57.)

[Atlas 2009] Scott W Atlas. Magnetic resonance imaging of the brain and spine, volume 1. Lippincott Williams & Wilkins, 2009. (Cited on page 56.)

[Baltzer 2013] Pascal A T Baltzer and Matthias Dietzel. *Breast lesions: diagnosis by using proton MR spectroscopy at 1.5 and 3.0 T–systematic review and meta-analysis.* Radiology, vol. 267, no. 3, pages 735–46, 2013. (Cited on pages 1 and 53.)

[Barker 1996] Fred G. Barker, Richard L. Davis, Susan M. Chang and Michael D. Prados. *Necrosis as a prognostic factor in glioblastoma multiforme*. Cancer, vol. 77, no. 6, pages 1161–1166, 1996. (Cited on page 167.)

[Baum 1985] J. Baum, R. Tycko and A. Pines. *Broadband and adiabatic inversion of a two-level system by phase-modulated pulses*. Physical Review A, vol. 32, no. 6, pages 3435–3447, 1985. (Cited on page 15.)

[Baumert 2006] Brigitta G. Baumert, Isabelle Rutten, Cary Dehing-Oberije, Albert Twijnstra, Miranda J.M. Dirx, Ria M.T.L. Debougnoux-Huppertz, Philippe Lambin and Bela Kubat. *A pathology-based substrate for target definition in radiosurgery of brain metastases*. International Journal of Radiation Oncology Biology Physics, vol. 66, no. 1, pages 187–194, 2006. (Cited on pages 60, 61 and 140.)

[Beer 2015] R De Beer and D Van Ormondt. *Java / JNI / C / Fortran based HSVD / HLSVD custom plugins for the j MRUI software system : Development , installation and usage*. no. L, pages 1–10, 2015. (Cited on page 35.)

[Bhattacharyya 2007] P. K. Bhattacharyya, M. J. Lowe and M. D. Phillips. *Spectral quality control in motion-corrupted single-voxel J-difference editing scans: An interleaved navigator approach*. Magnetic Resonance in Medicine, vol. 58, no. 4, pages 808–812, 2007. (Cited on page 93.)

[Bieniosek 2015] Matthew F. Bieniosek, Brian J. Lee and Craig S. Levin. *Technical Note: Characterization of custom 3D printed multimodality imaging phantoms*. Medical Physics, vol. 42, no. 10, pages 5913–5918, 2015. (Cited on page 78.)

[Bloch 1946] F. Bloch, W. W. Hansen and M. Packard. *The nuclear induction experiment*. Physical Review, vol. 70, no. 7-8, pages 474–485, 1946. (Cited on page 7.)

[Bodenhausen 1977] Geoffrey Bodenhausen, Ray Freeman and David L Turner. *Suppression of artifacts in two-dimensional J spectroscopy*. Journal of Magnetic Resonance (1969), vol. 27, no. 3, pages 511–514, 1977. (Cited on page 88.)

[Boesch 1999] Chris Boesch, Jacques Décombaz, Johannes Slotboom and Roland Kreis. *Observation of intramyocellular lipids by means of 1H magnetic resonance spectroscopy*. Proceedings of the Nutrition Society, vol. 58, no. 4, pages 841–850, 1999. (Cited on pages 1 and 53.)

[Boesch 2006] Chris Boesch, Juergen Machann, Peter Vermathen and Fritz Schick. *Role of proton MR for the study of muscle lipid metabolism*. NMR in Biomedicine, vol. 19, no. 7, pages 968–988, nov 2006. (Cited on pages 1 and 53.)

[Boichot 2006] C Boichot, P M Walker, C Durand, M Grimaldi, S Chapuis, J B Gouyon and F Brunotte. *Term neonate prognoses after perinatal asphyxia: contributions of MR imaging, MR spectroscopy, relaxation times, and apparent diffusion coefficients*. Radiology, vol. 239, no. 3, pages 839–848, 2006. (Cited on pages 1 and 53.)

[Bolan 2004] PJAW Bolan, P G Henry and M Garwood. *Feasibility of computer-intensive methods for estimating the variance of spectral fitting parameters*. In Proceedings of the 12th annual meeting ISMRM (abstract 304). Kyoto, 2004. (Cited on page 87.)

[Bolliger 2013] Christine S. Bolliger, Chris Boesch and Roland Kreis. *On the use of Cramér-Rao minimum variance bounds for the design of magnetic resonance spectroscopy experiments*. NeuroImage, vol. 83, pages 1031–1040, 2013. (Cited on pages 87 and 119.)

[Bottomley 1987] P a Bottomley. *Spatial localization in NMR spectroscopy in vivo*. Annals of the New York Academy of Sciences, vol. 508, pages 333–348, 1987. (Cited on page 26.)

[Bovée 1995] W. M M J Bovée, S. F. Keevil, M. O. Leach and F. Podo. *Quality assessment in in vivo NMR spectroscopy: II. A protocol for quality assessment. EEC concerted research project*. Magnetic Resonance Imaging, vol. 13, no. 1, pages 123–129, 1995. (Cited on pages 70, 71, 72, 73, 74 and 75.)

[Bovée 1998] W. Bovée, R. Canese, M. Decorps, E. Forssell-Aronsson, Y. Le Fur, F. Howe, O. Karlsen, A. Knijn, G. Kontaxis, H. Kgel, M. McLean, F. Podo, J. Slotboom, B. Vikhoff and A. Ziegler. *Absolute metabolite quantification by in vivo NMR spectroscopy: IV. Multicentre trial on MRSI localisation tests*. Magnetic Resonance Imaging, vol. 16, no. 9, pages 1113–1125, 1998. (Cited on pages 70, 77 and 188.)

[Breiman 1994] Leo Breiman. *Bagging predictors: Technical Report No. 421*. Machine Learning, vol. 140, no. 2, page 19, 1994. (Cited on pages 45 and 177.)

[Breiman 2001] Leo Breiman. *Random forests*. Machine Learning, vol. 45, no. 1, pages 5–32, 2001. (Cited on pages 44, 45, 82, 106, 121, 173, 176, 177 and 201.)

[Burger 1987] P C Burger and S B Green. *Patient age, histologic features, and length of survival in patients with glioblastoma multiforme*. Cancer, vol. 59, no. 9, pages 1617–1625, 1987. (Cited on page 167.)

[Burtscher 1999] Isabella M; Edvin Johansson; Stig Holtås; Freddy Ståhlberg Burtscher. *Quality Assessment of Localization Technique Performance in Small Volume in Vivo 1 H Mr Spectroscopy*. Magnetic resonance imaging, vol. 17, no. 10, pages 1511–1519, 1999. (Cited on page 76.)

[Cabanes 2001] E Cabanes, S Confort-Gouny, Y Le Fur, G Simond and P J Cozzone. *Optimization of residual water signal removal by HLSVD on simulated short echo time proton MR spectra of the human brain*. Journal of magnetic resonance (San Diego, Calif. : 1997), vol. 150, no. 2, pages 116–125, 2001. (Cited on page 35.)

[Canavagh 2007] J. Canavagh, W. J. Fairbrother, A. G. Palmer II, M. Rance and N. J. Skelton. *Theoretical Description of NMR Spectroscopy*. Protein NMR Spectroscopy 2nd Edit, 2007. (Cited on page 24.)

[Castillo 2000] M. Castillo, J. K. Smith and L. Kwock. *Correlation of myo-inositol levels and grading of cerebral astrocytomas*. American Journal of Neuroradiology, vol. 21, no. 9, pages 1645–1649, 2000. (Cited on page 57.)

[Cavassila 2001] S. Cavassila, S. Deval, C. Huegen, D. Van Ormondt and D. Graveron-Demilly. *Cramér-Rao bounds: An evaluation tool for quantitation*. NMR in Biomedicine, vol. 14, no. 4, pages 278–283, 2001. (Cited on pages 86 and 119.)

[Cen 2014] Dongzhi Cen and Li Xu. *Differential diagnosis between malignant and benign breast lesions using single-voxel proton MRS: A meta-analysis*. Journal of Cancer Research and Clinical Oncology, vol. 140, no. 6, pages 993–1001, 2014. (Cited on pages 1 and 53.)

[Choi 2011] Changho Choi, Sandeep K. Ganji, Ralph J. Deberardinis, Ivan E. Dimitrov, Juan M. Pascual, Robert Bachoo, Bruce E. Mickey, Craig R. Malloy and Elizabeth A. Maher. *Measurement of glycine in the human brain in vivo by 1H-MRS at 3 T: Application in brain tumors*. Magnetic Resonance in Medicine, vol. 66, no. 3, pages 609–618, 2011. (Cited on page 55.)

[Choi 2012] Changho Choi, Sandeep K Ganji, Ralph J DeBerardinis, Kimmo J Hatanpaa, Dinesh Rakheja, Zoltan Kovacs, Xiao-Li Yang, Tomoyuki Mashimo, Jack M Raisanen, Isaac Marin-Valencia, Juan M Pascual, Christopher J Madden, Bruce E Mickey, Craig R Malloy, Robert M Bachoo and Elizabeth A Maher. *2-hydroxyglutarate detection by magnetic resonance spectroscopy in IDH-mutated patients with gliomas.* Nature medicine, vol. 18, no. 4, pages 624–9, 2012. (Cited on pages 57 and 69.)

[Choi 2013] Changho Choi, Sandeep Ganji, Keith Hulsey, Akshay Madan, Zoltan Kovacs, Ivan Dimitrov, Song Zhang, Kumar Pichumani, Dianne Mendelsohn, Bruce Mickey, Craig Malloy, Robert Bachoo, Ralph Deberardinis and Elizabeth Maher. *A comparative study of short- and long-TE 1H MRS at 3 T for in vivo detection of 2-hydroxyglutarate in brain tumors.* NMR in biomedicine, vol. 26, no. 10, pages 1242–50, 2013. (Cited on page 57.)

[Claes 2007] An Claes, Albert J. Idema and Pieter Wesseling. *Diffuse glioma growth: A guerilla war*. Acta Neuropathologica, vol. 114, no. 5, pages 443–458, 2007. (Cited on page 171.)

[Cohn 1996] David A. Cohn, Zoubin Ghahramani and Michael I. Jordan. *Active learning with statistical models*. Journal of Artificial Intelligence Research, vol. 4, pages 129–145, 1996. (Cited on page 120.)

[Coman 2009] Daniel Coman, Hubert K. Trubel, Robert E. Rycyna and Fahmeed Hyder. *Brain temperature and pH measured by 1 H chemical shift imaging of a thulium agent*. NMR in Biomedicine, vol. 22, no. 2, pages 229–239, 2009. (Cited on page 38.)

[Conolly 1989a] Steven Conolly, Dwight Nishimura and Albert Macovski. *A selective adiabatic spin-echo pulse*. Journal of Magnetic Resonance (1969), vol. 83, no. 2, pages 324–334, 1989. (Cited on page 92.)

[Conolly 1989b] Steven Conolly, Dwight Nishimura and Albert Macovski. *Sweep-diagram analysis of selective adiabatic pulses*. Journal of Magnetic Resonance (1969), vol. 83, no. 3, pages 549–564, 1989. (Cited on page 92.)

[Cordova 2016] James S. Cordova, Hui Kuo G Shu, Zhongxing Liang, Saumya S. Gurbani, Lee A D Cooper, Chad A. Holder, Jeffrey J. Olson, Brad Kairdolf, Eduard Schreibmann, Stewart G. Neill, Constantinos G. Hadjipanayis and Hyunsuk Shim. *Whole-brain spectroscopic MRI biomarkers identify infiltrating margins in glioblastoma patients*. Neuro-Oncology, vol. 18, no. 8, pages 1180–1189, 2016. (Cited on pages 62, 143, 147, 151, 152, 158, 166, 167, 172, 183, 188 and 189.)

[Costello 1999] L C Costello, R B Franklin and P Narayan. *Citrate in the diagnosis of prostate cancer.* The Prostate, vol. 38, no. 3, pages 237–45, 1999. (Cited on pages 1 and 53.)

[Cramér 1946] Harald Cramér. *Mathematical models of statistics*. Princeton Mathematical Series. Princeton University Press, Princteon, NJ, 1946. (Cited on page 83.)

[Criminisi 2013] A. Criminisi and J. Shotton. *Classification Forests*. In Decision Forests for Computer Vision and Medical Image Analysis, pages 25–45. 2013. (Cited on pages 46 and 106.)

[Croteau 2001] David Croteau, Lisa Scarpace, David Hearshen, Jorge Gutierrez, James L. Fisher, Jack P. Rock and Tom Mikkelsen. *Correlation between magnetic resonance spectroscopy imaging and image-guided biopsies: Semiquantitative and qualitative histopathological analyses of patients with untreated glioma*. Neurosurgery, vol. 49, no. 4, pages 823–829, 2001. (Cited on page 143.)

[Cudalbu 2012] Cristina Cudalbu, Vladimir Mlynarik and Rolf Gruetter. *Handling macromolecule signals in the quantification of the neurochemical profile*. Journal of Alzheimer's Disease, vol. 31, no. SUPPL. 3, 2012. (Cited on page 56.)

[Dang 2009] Lenny Dang, David W. White, Stefan Gross, Bryson D. Bennett, Mark A. Bittinger, Edward M. Driggers, Valeria R. Fantin, Hyun Gyung Jang, Shengfang Jin, Marie C. Keenan, Kevin M. Marks, Robert M. Prins, Patrick S. Ward, Katharine E. Yen, Linda M. Liau, Joshua D. Rabinowitz, Lewis C. Cantley, Craig B. Thompson, Matthew G. Vander Heiden and Shinsan M. Su. *Cancer-associated IDH1 mutations produce 2-hydroxyglutarate*. Nature, vol. 462, no. 7274, pages 739–744, 2009. (Cited on pages 57 and 69.)

[de Beer 1992] R de Beer, D van Ormondt and W W F Pijnappel. *Quantification of 1-D and 2-D magnetic resonance time domain signals*. Pure & Appl. Chem., vol. 64, no. 6, pages 815–823, 1992. (Cited on pages 34, 70 and 73.)

[De Beer 1998a] R. De Beer, B. Barbiroli, G. Gobbi, A. Knijn, H. Kgel, K. W. Langenberger, I. Tkac and S. Topp. *Absolute metabolite quantification by in vivo NMR spectroscopy: III. Multicentre 1H MRS of the human brain addressed by one and the same data- analysis protocol*. Magnetic Resonance Imaging, vol. 16, no. 9, pages 1107–1111, 1998. (Cited on page 70.)

[De Beer 1998b] R. De Beer, A. Van Den Boogaart, E. Cady, D. Graveron-Demilly, A. Knijn, K. W. Langenberger, J. C. Lindon, A. Ohlhoff, H. Serrafi and M. Wylezinska-Arridge. *Absolute metabolite quantification by in vivo NMR spectroscopy: V. Multicentre quantitative data analysis trial on the overlapping background problem*. Magnetic Resonance Imaging, vol. 16, no. 9, pages 1127–1137, 1998. (Cited on page 70.)

[De Edelenyi 2000] Fabien Szabo De Edelenyi, Christophe Rubin, François Esteve and Sylvie Grand. *A new approach for analyzing proton magnetic resonance spectroscopic images of brain tumors: nosologic images*. Nature Medicine, vol. 6, no. 11, pages 1287–1289, 2000. (Cited on page 172.)

[De Graaf 1990] A. A. De Graaf, J. E. Van Dijk and W. M. M. J. Bovée. *QUALITY: quantification improvement by converting lineshapes to the lorentzian type*. Magnetic Resonance in Medicine, vol. 13, no. 3, pages 343–357, mar 1990. (Cited on page 88.)

[Debarr 2009] Dave Debarr and Harry Wechsler. *Spam Detection using Clustering , Random Forests , and Active Learning*. Sixth Conference on Email and Anti-Spam, pages 1–3, 2009. (Cited on page 120.)

[DeBerardinis 2010] R J DeBerardinis and T Cheng. *Q's next: the diverse functions of glutamine in metabolism, cell biology and cancer*. Oncogene, vol. 29, no. 3, pages 313–324, 2010. (Cited on pages 55 and 189.)

[den Boogaart 1994] A den Boogaart, D Van Ormondt, W W F Pijnappel, R De Beer and M Ala-Korpela. *Removal of the water resonance from 1H magnetic resonance spectra*. Mathematics in signal processing, vol. 3, pages 175–195, 1994. (Cited on pages 149 and 174.)

[Ding 2010] Chris Ding, Tao Li and Michael I. Jordan. *Convex and semi-nonnegative matrix factorizations*. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 32, no. 1, pages 45–55, 2010. (Cited on pages 44, 49 and 50.)

[Dowling 2001] C. Dowling, A. W. Bollen, S. M. Noworolski, M. W. McDermott, N. M. Barbaro, M. R. Day, R. G. Henry, S. M. Chang, W. P. Dillon, S. J. Nelson and D. B. Vigneron. *Preoperative proton MR spectroscopic imaging of brain tumors: Correlation with histopathologic analysis of resection specimens*. American Journal of Neuroradiology, vol. 22, no. 4, pages 604–612, 2001. (Cited on pages 1 and 53.)

[Du 2004] Shuyan Du, Paul Sajda, Xiangling Mao and Dikoma Shungu. *Du S, Sajda P, Mao X, Shungu D. Multiresolution hierarchical blind recovery of biochemical markers of brain cancer in MRSI*. IEEE Biomedical Imaging: Nano to Macro, pages 233–236, 2004. (Cited on pages 172 and 189.)

[Du 2008] Shuyan Du, Xiangling Mao, Paul Sajda and Dikoma C. Shungu. *Automated tissue segmentation and blind recovery of1H MRS imaging spectral patterns of normal and diseased human brain*. NMR in Biomedicine, vol. 21, no. 1, pages 33–41, jan 2008. (Cited on pages 172 and 189.)

[Efron 1981] Bradley Efron. *Nonparametric Standard Errors and Confidence Intervals*. The Canadian Journal of Statistics / La Revue Canadienne de Statistique, vol. 9, no. 2, pages 139–158, 1981. (Cited on page 153.)

[Einstein 2012] Douglas B. Einstein, Barry Wessels, Barbara Bangert, Pingfu Fu, A. Dennis Nelson, Mark Cohen, Stephen Sagar, Jonathan Lewin, Andrew Sloan, Yiran Zheng, Jordonna Williams, Valdir Colussi, Robert Vinkler and Robert Maciunas. *Phase II trial of radiosurgery to magnetic resonance spectroscopy-defined high-risk tumor volumes in patients with glioblastoma multiforme*. International Journal of Radiation Oncology Biology Physics, vol. 84, no. 3, pages 668–674, 2012. (Cited on pages 62, 167 and 172.)

[Ernst 1996] Thomas Ernst and Linda Chang. *Elimination of artifacts in short echo time 1H MR spectroscopy of the frontal lobe*. Magnetic Resonance in Medicine, vol. 36, no. 3, pages 462–468, 1996. (Cited on pages 88 and 89.)

[Evans 1999] James Robert Evans and William M Lindsay. *The management and control of quality*. 1999. (Cited on page 67.)

[Evaristi 2016] Maria Francesca Evaristi, Celine Caubere, Romain Harmancey, Franck Desmoulin, William Frank Peacock, Matthieu Berry, Annie Turkieh, Manon Barutaut, Michel Galinier, Camille Dambrin, Carlo Polidori, Cristina Miceli, Bernard Chamontin, Francois Koukoui, Jerome Roncalli, Pierre Massabuau, Fatima Smih and Philippe Rouet. *Increased mean aliphatic lipid chain length in left ventricular hypertrophy secondary to arterial hypertension: A cross-sectional study.* Medicine, vol. 95, no. 46, page e4965, 2016. (Cited on page 56.)

[Frahm 1987] Jens Frahm, Klaus-Dietmar Merboldt and Wolfgang Hänicke. *Localized proton spectroscopy using stimulated echoes*. Journal of Magnetic Resonance (1969), vol. 72, no. 3, pages 502–508, 1987. (Cited on page 27.)

[Ganslandt 2005] Oliver Ganslandt, Andreas Stadlbauer, Rudolf Fahlbusch, Kyosuke Kamada, Rolf Buslei, Ingmar Blumcke, Ewald Moser and Christopher Nimsky. *Proton magnetic resonance spectroscopic imaging integrated into image-guided surgery: Correlation to standard magnetic resonance imaging and tumor cell density*. Neurosurgery, vol. 56, no. 4 SUPPL., pages 291–298, 2005. (Cited on pages 62, 172 and 183.)

[Gaparovic 2016] Chuck Gaparovic. *Adjusting Concentrations for Tissue Composition in Brain : Assumptions & Conundrums*. In ISMRM Workshop on MR Spectroscopy, Lake Constance, 2016. (Cited on page 42.)

[García-Figueiras 2016] Roberto García-Figueiras, Sandra Baleato-González, Anwar R. Padhani, Laura Oleaga, Joan C. Vilanova, Antonio Luna and Juan Carlos Cobas Gómez. *Proton magnetic resonance spectroscopy in oncology: The fingerprints of cancer?* Diagnostic and Interventional Radiology, vol. 22, no. 1, pages 75–89, 2016. (Cited on pages 1, 2, 53, 55, 57 and 148.)

[Garwood 2001] Michael Garwood and Lance DelaBarre. *The Return of the Frequency Sweep: Designing Adiabatic Pulses for Contemporary NMR*. Journal of Magnetic Resonance, vol. 153, no. 2, pages 155–177, 2001. (Cited on page 30.)

[Goelman 2007] Gadi Goelman, Songtao Liu, Roman Fleysher, Lazar Fleysher, Robert I. Grossman and Oded Gonen. *Chemical-shift artifact reduction in hadamard-encoded MR spectroscopic imaging at high (3T and 7T) magnetic fields*. Magnetic Resonance in Medicine, vol. 58, no. 1, pages 167–173, 2007. (Cited on page 90.)

[Gorter 1942] C. J. Gorter and L. J. F. Broer. *Negative Result of an Attempt to Observe Nuclear Magnetic Resonance in Solids*. Physica, vol. 9, page 591, 1942. (Cited on page 7.)

[Govindaraju 2000] Varanavasi Govindaraju, Karl Young and Andrew A. Maudsley. *Proton NMR chemical shifts and coupling constants for brain metabolites*. NMR in Biomedicine, vol. 13, no. 3, pages 129–153, 2000. (Cited on pages 21, 26, 53, 55, 57 and 188.)

[Greenhaff 2001] Paul L Greenhaff. *The creatine-phosphocreatine system: there's more than one song in its repertoire*. Journal of physiology, vol. 537, no. 3, page 657, 2001. (Cited on page 55.)

[Guo 2012] Jun Guo, Chengjun Yao, Hong Chen, Dongxiao Zhuang, Weijun Tang, Guang Ren, Yin Wang, Jinsong Wu, Fengping Huang and Liangfu Zhou. *The relationship between cho/naa and glioma metabolism: Implementation for margin delineation of cerebral gliomas*. Acta Neurochirurgica, vol. 154, no. 8, pages 1361–1370, 2012. (Cited on pages 147, 151, 152, 167, 172 and 183.)

[Haase 1985] A Haase, J Frahm, W Hänicke and D Matthaei. *1H NMR chemical shift selective (CHESS) imaging*. Physics in medicine and biology, vol. 30, no. 4, pages 341–344, 1985. (Cited on page 94.)

[Hájek 2000] Milan Hájek, Martin Burian and Monika Dezortová. *Application of LCModel for quality control and quantitative in vivo 1H MR spectroscopy by short echo time STEAM sequence*. Magnetic Resonance Materials in Physics, Biology and Medicine, vol. 10, no. 1, pages 6–17, 2000. (Cited on page 68.)

[Hall 2009] Mark Hall, Eibe Frank, Geoffrey Holmes, Bernhard Pfahringer, Peter Reutemann and Ian H Witten. *The WEKA data mining software: an update*. ACM SIGKDD explorations newsletter, vol. 11, no. 1, pages 10–18, 2009. (Cited on page 151.)

[Harting 2015] Inga Harting, Nikolas Boy, Jana Heringer, Angelika Seitz, Martin Bendszus, Petra J W Pouwels and Stefan Kölker. *1H-MRS in glutaric aciduria type 1: impact of biochemical phenotype and age on the cerebral accumulation of neurotoxic metabolites*. Journal of Inherited Metabolic Disease, vol. 38, no. 5, pages 829–838, 2015. (Cited on pages 1 and 53.)

[Hastie 2001] Trevor Hastie, Robert Tibshirani and Jerome Friedman. The elements of statistical learning, volume 1. New York: Springer series in statistics, 2001. (Cited on pages 45, 46, 48, 106 and 176.)

[Hattingen 2008] Elke Hattingen, Peter Raab, Kea Franz, Friedhelm E. Zanella, Heinrich Lanfermann and Ulrich Pilatus. *Myo-Inositol: a marker of reactive astrogliosis in glial tumors?* NMR in Biomedicine, vol. 21, no. 3, pages 233–241, apr 2008. (Cited on page 57.)

[Heikal 2008] A A Heikal, K Wachowicz, S D Thomas and B G Fallone. *A phantom to assess the accuracy of tumor delineation using MRSI*. Radiology and Oncology, vol. 42, no. 4, pages 232–239, 2008. (Cited on pages 76, 77 and 188.)

[Hermann 2008] Elvis J. Hermann, Elke Hattingen, Joachim K. Krauss, Gerhard Marquardt, Ulrich Pilatus, Kea Franz, Matthias Setzer, Thomas Gasser, Dominique S. Tews, Friedhelm E. Zanella, Volker Seifert and Heinrich Lanfermann. *Stereotactic biopsy in gliomas guided by 3-tesla 1H-chemical- shift imaging of choline*. Stereotactic and Functional Neurosurgery, vol. 86, no. 5, pages 300–307, 2008. (Cited on page 167.)

[Herminghaus 2003] Sebastian Herminghaus, Thomas Dierks, Ulrich Pilatus, Walter Möller-Hartmann, Jörg Wittsack, Gerhard Marquardt, Christoph Labisch, Heinrich Lanfermann, Wolfgang Schlote and Friedhelm E. Zanella. *Determination of histopathological tumor grade in neuroepithelial brain tumors by using spectral pattern analysis of in vivo spectroscopic data*. Journal of Neurosurgery, vol. 98, no. 1, pages 74–81, 2003. (Cited on page 61.)

[Hess 2011] Aaron T. Hess, M. Dylan Tisdall, Ovidiu C. Andronesi, Ernesta M. Meintjes and André J W Van Der Kouwe. *Real-time motion and B0 corrected single voxel spectroscopy using volumetric navigators*. Magnetic Resonance in Medicine, vol. 66, no. 2, pages 314–323, 2011. (Cited on page 93.)

[Hess 2013] Aaron T. Hess, Andre J W Van Der Kouwe, Kenneth K. Mbugua, Barbara Laughton and Ernesta M. Meintjes. *Quality of 186 child brain spectra using motion and b0 shim navigated single voxel spectroscopy*. Journal of Magnetic Resonance Imaging, vol. 40, no. 4, pages 958–965, 2013. (Cited on page 93.)

[Hock 2013] Andreas Hock, Erin L. MacMillan, Alexander Fuchs, Roland Kreis, Peter Boesiger, Spyros S. Kollias and Anke Henning. *Non-water-suppressed proton MR spectroscopy improves spectral quality in the human spinal cord*. Magnetic Resonance in Medicine, vol. 69, no. 5, pages 1253–1260, 2013. (Cited on page 92.)

[Hollingworth 2006] W Hollingworth, L S Medina, R E Lenkinski, D K Shibata, B Bernal, D Zurakowski, B Comstock and J G Jarvik. *A systematic literature review of magnetic resonance spectroscopy for the characterization of brain tumors*. AJNR American journal of neuroradiology, vol. 27, no. 7, pages 1404–1411, 2006. (Cited on pages 58, 61, 147, 151 and 172.)

[Horská 2010] Alena Horská and Peter B. Barker. *Imaging of brain tumors: MR spectroscopy and metabolic imaging*, 2010. (Cited on pages 58, 151, 152 and 167.)

[Hourani 2008] R. Hourani, L. J. Brant, T. Rizk, J. D. Weingart, P. B. Barker and Alena Horská. *Can proton MR spectroscopic and perfusion imaging differentiate between neoplastic and nonneoplastic brain lesions in adults?* American Journal of Neuroradiology, vol. 29, no. 2, pages 366–372, 2008. (Cited on page 58.)

[Howe 1995] F. A. Howe, R. Canese, F. Podo, B. Vikhoff, J. Slotboom, J. R. Griffiths, O. Henriksen and W. M M J Bovée. *Quality assessment in in vivo NMR spectroscopy: V. Multicentre evaluation of prototype test objects and protocols for performance assessment in small bore MRS equipment*. Magnetic Resonance Imaging, vol. 13, no. 1, pages 159–167, 1995. (Cited on pages 70, 73 and 75.)

[Howe 2003] F A Howe, S J Barton, S A Cudlip, M Stubbs, D E Saunders, M Murphy, P Wilkins, K S Opstad, V L Doyle, M A McLean, B A Bell and J R Griffiths. *Metabolic profiles of human brain tumors using quantitative in vivo 1H magnetic resonance spectroscopy.* Magnetic resonance in medicine : official journal of the Society of Magnetic Resonance in Medicine / Society of Magnetic Resonance in Medicine, vol. 49, pages 223–232, 2003. (Cited on pages 1 and 53.)

[Huber 2017] T. Huber, G. Alber, S. Bette, T. Boeckh-Behrens, J. Gempt, F. Ringel, E. Alberts, C. Zimmer and J. S. Bauer. *Reliability of Semi-Automated Segmentations in Glioblastoma*. Clinical Neuroradiology, vol. 27, no. 2, pages 153–161, 2017. (Cited on page 190.)

[Ibrahim 2007] Tamer S. Ibrahim and Lin Tang. *Insight into RF power requirements and B1 field homogeneity for human MRI via rigorous FDTD approach*. Journal of Magnetic Resonance Imaging, vol. 25, no. 6, pages 1235–1247, 2007. (Cited on page 28.)

[Jain 2017] Saurabh Jain, Diana M. Sima, Faezeh Sanaei Nezhad, Gilbert Hangel, Wolfgang Bogner, Stephen Williams, Sabine Van Huffel, Frederik Maes and Dirk Smeets. *Patch-Based Super-Resolution of MR Spectroscopic Images: Application to Multiple Sclerosis*. Frontiers in Neuroscience, vol. 11, 2017. (Cited on page 188.)

[Jiru 2006] F. Jiru, A. Skoch, U. Klose, W. Grodd and M. Hajek. *Error images for spectroscopic imaging by LCModel using Cramer-Rao bounds*. Magnetic Resonance Materials in Physics, Biology and Medicine, vol. 19, no. 1, pages 1–14, 2006. (Cited on pages 86, 87, 100 and 119.)

[Joanes 1998] D. N. Joanes and C. A. Gill. *Comparing measures of sample skewness and kurtosis*. Journal of the Royal Statistical Society: Series D (The Statistician), vol. 47, no. 1, pages 183–189, 1998. (Cited on page 83.)

[Julià-Sapé 2006] Margarida Julià-Sapé, Dionisio Acosta, Mariola Mier, Carles Arús and Des Watson. *A multi-centre, web-accessible and quality control-checked database of in vivo MR spectra of brain tumour patients*. Magnetic Resonance Materials in Physics, Biology and Medicine, vol. 19, no. 1, pages 22–33, 2006. (Cited on page 75.)

[Julià-Sapé 2012] Margarida Julià-Sapé, Miguel Lurgi, Mariola Mier, Francesc Estanyol, Xavier Rafael, Ana Paula Candiota, Anna Barceló, Alina García, M Carmen Martínez-Bisbal, Rubén Ferrer-Luna, Ángel Moreno-Torres, Bernardo Celda and Carles Arús. *Strategies for annotation and curation of translational databases: the eTUMOUR project.* Database : the journal of biological databases and curation, vol. 2012, page bas035, 2012. (Cited on page 82.)

[Kaiser 2007] Lana G. Kaiser, Karl Young and Gerald B. Matson. *Elimination of spatial interference in PRESS-localized editing spectroscopy*. Magnetic Resonance in Medicine, vol. 58, no. 4, pages 813–818, 2007. (Cited on pages 90 and 92.)

[Kasten 2016] Jeffrey A Kasten, Thomas Vetterli, Francois Lazeyras and Dimitri de Ville. *3D-Printed Shepp-Logan Phantom as a Real-World Benchmark for MRI*. Magnetic Resonance in Medicine, vol. 75, no. 1, pages 287–294, 2016. (Cited on page 78.)

[Keeler 2004] James Keeler. Understanding NMR spectroscopy. Barcelona, 2004. (Cited on pages 23 and 25.)

[Keevil 1995] S. F. Keevil, B. Barbiroli, D. J. Collins, E. R. Danielsen, J. Hennig, O. Henriksen, M. O. Leach, R. Longo, M. Lowry, C. Moore, E. Moser, C. Segebarth, W. M M J Bovée and F. Podo. *Quality assessment in in vivo NMR spectroscopy: IV. A multicentre trial of test objects and protocols for performance assessment in clinical NMR spectroscopy*. Magnetic Resonance Imaging, vol. 13, no. 1, pages 139–157, 1995. (Cited on pages 70, 73 and 74.)

[Keevil 1998] S. F. Keevil, B. Barbiroli, J. C W Brooks, E. B. Cady, R. Canese, P. Carlier, D. J. Collins, P. Gilligan, G. Gobbi, J. Hennig, H. Kügel, M. O. Leach, D. Metzler, V. Mlynárik, E. Moser, M. C. Newbold, G. S. Payne, P. Ring, J. N. Roberts, I. J. Rowland, T. Thiel, I. Tkác, S. Topp, H. J. Wittsack, M. Wylezinska, P. Zaniol, O. Henriksen and F. Podo. *Absolute metabolite quantification by in vivo NMR spectroscopy: II. A multicentre trial of protocols for in vivo localised proton studies of human brain*. Magnetic Resonance Imaging, vol. 16, no. 9, pages 1093–1106, 1998. (Cited on page 70.)

[Kreis 1993] R. Kreis, T. Ernst and B.D. Ross. *Absolute Quantitation of Water and Metabolites in the Human Brain. II. Metabolite Concentrations*. Journal of Magnetic Resonance, Series B, vol. 102, no. 1, pages 9–19, aug 1993. (Cited on page 80.)

[Kreis 1997] R. Kreis. *Quantitative localized 1H MR spectroscopy for clinical use*. Progress in Nuclear Magnetic Resonance Spectroscopy, vol. 31, no. 97, pages 155–195, 1997. (Cited on page 57.)

[Kreis 2004] Roland Kreis. *Issues of spectral quality in clinical 1H-magnetic resonance spectroscopy and a gallery of artifacts*. NMR in Biomedicine, vol. 17, no. 6, pages 361–381, 2004. (Cited on pages 79, 80, 92, 99, 103, 119 and 188.)

[Kreis 2012] Roland Kreis. *Clinical MRS: Promise, Potential, Power & Pitfalls*. In Proceedings of the 20th Annual Meeting of ISMRM, 2012. (Cited on pages 1 and 2.)

[Kreis 2015] Roland Kreis. *The trouble with quality filtering based on relative Cramér-Rao lower bounds*. Magnetic Resonance in Medicine, vol. 00, 2015. (Cited on pages 86, 100 and 119.)

[Kuhn 2013] Max Kuhn and Kjell Johnson. Applied predictive modeling. 2013. (Cited on page 101.)

[Kutsuna 2012] Natsumaro Kutsuna, Takumi Higaki, Sachihiro Matsunaga, Tomoshi Otsuki, Masayuki Yamaguchi, Hirofumi Fujii and Seiichiro Hasezawa. *Active learning framework with iterative clustering for bioimage classification*. Nature Communications, vol. 3, no. May, page 1032, 2012. (Cited on page 120.)

[Kyathanahally 2017] Sreenath P. Kyathanahally, Victor Mocioiu, Nuno Pedrosa de Barros, Johannes Slotboom, Alan J. Wright, Margarida Julià-Sapé, Carles Arús and Roland Kreis. *Quality of clinical brain tumor MR spectra judged by humans and machine learning tools*. Magnetic Resonance in Medicine, vol. 00, no. September, 2017. (Cited on page 202.)

[Lai 2005] Ping H Lai, Kun T Li, Shu S Hsu, Chia C Hsiao, Chi W Yip, S Ding, Lee R Yeh and Huay B Pan. *Pyogenic Brain Abscess : Findings from In Vivo 1 . 5-T and 11 . 7-T In Vitro Proton MR Spectroscopy*. Ajnr, no. February, pages 279–288, 2005. (Cited on pages 58 and 59.)

[Lam 2015] Fan Lam, Chao Ma, Bryan Clifford, Curtis L. Johnson and Zhi-Pei Liang. *High-resolution 1 H-MRSI of the brain using SPICE: Data acquisition and image reconstruction*. Magnetic Resonance in Medicine, vol. 00, no. September, pages n/a–n/a, 2015. (Cited on pages 81, 166 and 206.)

[Lange 2006] T. Lange, U. Dydak, T. P L Roberts, H. A. Rowley, M. Bjeljac and Peter Boesiger. *Pitfalls in lactate measurements at 3T*. American Journal of Neuroradiology, vol. 27, no. 4, pages 895–901, 2006. (Cited on page 90.)

[Larjavaara 2007] S. Larjavaara, R. Mantyla, T. Salminen, H. Haapasalo, J. Raitanen, J. Jaaskelainen and A. Auvinen. *Incidence of gliomas by anatomic location*. Neuro-Oncology, vol. 9, no. 3, pages 319–325, 2007. (Cited on page 156.)

[Law 2003] Meng Law, Stanley Yang, Hao Wang, James S Babb, Glyn Johnson, Soonmee Cha, Edmond A Knopp and David Zagzag. *Glioma grading: sensitivity, specificity, and predictive values of perfusion MR imaging and proton MR spectroscopic imaging compared with conventional MR imaging*. AJNR. American journal of neuroradiology, vol. 24, no. 10, pages 1989–98, 2003. (Cited on pages 147, 151, 152, 158 and 167.)

[Leach 1995] M. O. Leach, D. J. Collins, S. Keevil, I. Rowland, M. A. Smith, O. Henriksen, W. M M J Bovée and F. Podo. *Quality assessment in in vivo NMR spectroscopy: III. Clinical test objects: Design, construction, and solutions*. Magnetic Resonance Imaging, vol. 13, no. 1, pages 131–137, 1995. (Cited on pages 70, 73 and 74.)

[Lewis 1994] David D. Lewis and William a. Gale. *A Sequential Algorithm For Training Text Classifiers*. Proceedings of the 17th International Conference on Research and Development in Information Retrieval (SIGIR'94), pages 3–12, 1994. (Cited on pages 82 and 120.)

[Li 2013a] Yuqian Li, Diana M. Sima, Sofie Van Cauter, Anca R. Croitor Sava, Uwe Himmelreich, Yiming Pi and Sabine Van Huffel. *Hierarchical non-negative matrix factorization (hNMF): A tissue pattern differentiation method for glioblastoma multiforme diagnosis using MRSI*. NMR in Biomedicine, vol. 26, no. 3, pages 307–319, 2013. (Cited on pages 148, 172 and 189.)

[Li 2013b] Yuqian Li, Diana M. Sima, Sofie Van Cauter, Uwe Himmelreich, Anca R Croitor Sava, Yiming Pi, Yipeng Liu and Sabine Van Huffel. *Unsupervised nosologic imaging for glioma diagnosis*. IEEE Transactions on Biomedical Engineering, vol. 60, no. 6, pages 1760–1763, 2013. (Cited on pages 172 and 189.)

[Liaw 2002] a Liaw and M Wiener. *Classification and Regression by randomForest*. R news, vol. 2, no. December, pages 18–22, 2002. (Cited on pages 44, 45, 48, 82, 83, 101, 106, 121, 173, 176 and 205.)

[Lichy 2004] Matthias Philipp Lichy, M. Henze, C. Plathow, P. Bachert, H. U. Kauczor and H. P. Schlemmer. *Metabolische bildgebung zur verlaufskontrolle stereotaktisch bestrahlter gliome - Wertigkeit der 1H-MR-spektroskopie im vergleich zur FDG-PET und IMT-SPECT*. RoFo Fortschritte auf dem Gebiet der Rontgenstrahlen und der Bildgebenden Verfahren, vol. 176, no. 8, pages 1114–1121, 2004. (Cited on pages 147 and 151.)

[Lopez 2017] Christopher J. Lopez, Natalya Nagornaya, Nestor A. Parra, Deukwoo Kwon, Fazilat Ishkanian, Arnold M. Markoe, Andrew Maudsley and Radka Stoyanova. *Association of Radiomics and Metabolic Tumor Volumes in Radiation Treatment of Glioblastoma Multiforme*. International Journal of Radiation Oncology Biology Physics, vol. 97, no. 3, pages 586–595, 2017. (Cited on pages 163 and 177.)

[Lu 2013] Deji Lu, Yu Sun and Suiren Wan. *Brain Tumor Classification Using Non-negative and Local Non-negative Matrix Factorization*. 2013 IEEE International Conference on Signal Processing, Communication and Computing (ICSPCC), pages 1–4, 2013. (Cited on page 172.)

[Luts 2009] Jan Luts, Teresa Laudadio, Albert J. Idema, Arjan W. Simonetti, Arend Heerschap, Dirk Vandermeulen, Johan A K Suykens and Sabine Van Huffel. *Nosologic imaging of the brain: Segmentation and classification using MRI and MRSI*. NMR in Biomedicine, vol. 22, no. 4, pages 374–390, 2009. (Cited on pages 148 and 172.)

[Ma 2016] Chao Ma, Fan Lam, Qiang Ning, Curtis L Johnson and Zhi-Pei Liang. *High-resolution (1) H-MRSI of the brain using short-TE SPICE.* Magnetic resonance in medicine, 2016. (Cited on pages 81, 166 and 206.)

[Mader 2001] I Mader, U Seeger, R Weissert, U Klose, T Naegele, A Melms and W Grodd. *Proton MR spectroscopy with metabolite-nulling reveals elevated macromolecules in acute multiple sclerosis.* Brain, vol. 124, pages 953–961, 2001. (Cited on page 14.)

[Maiora 2014] Josu Maiora, Borja Ayerdi and Manuel Graña. *Random forest active learning for AAA thrombus segmentation in computed tomography angiography images*. Neurocomputing, vol. 126, pages 71–77, 2014. (Cited on page 120.)

[Majós 2004] Carles Majós, Margarida Julià-Sapé, Juli Alonso, Marta Serrallonga, Carles Aguilera, Juan J Acebes, Carles Arús and Jaume Gili. *Brain tumor classification by proton MR spectroscopy: comparison of diagnostic accuracy at short and long TE.* AJNR. American journal of neuroradiology, vol. 25, no. 10, pages 1696–704, 2004. (Cited on pages 1 and 53.)

[Matsusue 2010] Eiji Matsusue, James R. Fink, Jason K. Rockhill, Toshihide Ogawa and Kenneth R. Maravilla. *Distinction between glioma progression and post-radiation change by combined physiologic MR imaging*. Neuroradiology, vol. 52, no. 4, pages 297–306, 2010. (Cited on pages 147 and 151.)

[Meier 2012] D. Meier, P. Boesiger and S. Kozerke. Magnetic Resonance Imaging in Medicine. University and ETH Zürich, Switzerland, Zürich, 2012. (Cited on page 11.)

[Meier 2016] Raphael Meier, Urspeter Knecht, Tina Loosli, Stefan Bauer, Johannes Slotboom, Roland Wiest and Mauricio Reyes. *Clinical Evaluation of a Fully-automatic Segmentation Method for Longitudinal Brain Tumor Volumetry.* Scientific reports, vol. 6, no. October 2015, page 23376, 2016. (Cited on pages 141, 142, 148, 150 and 174.)

[Meldrum 2000] B S Meldrum. *Glutamate as a neurotransmitter in the brain: review of physiology and pathology.* The Journal of nutrition, vol. 130, no. 4S Suppl, pages 1007S–15S, 2000. (Cited on page 55.)

[Menze 2008] Bjoern H. Menze, B. Michael Kelm, Marc André Weber, Peter Bachert and Fred A. Hamprecht. *Mimicking the human expert: Pattern recognition for an automated assessment of data quality in MR spectroscopic images*. Magnetic Resonance in Medicine, vol. 59, no. 6, pages 1457–1466, 2008. (Cited on pages 81, 82, 101, 104, 112, 113 and 120.)

[Menze 2009] Bjoern H Menze, B Michael Kelm, Ralf Masuch, Uwe Himmelreich, Peter Bachert, Wolfgang Petrich and Fred A Hamprecht. *A comparison of random forest and its Gini importance with standard chemometric methods for the feature selection and classification of spectral data*. BMC Bioinformatics, vol. 10, no. 1, page 213, 2009. (Cited on page 46.)

[Metwally 2014] Lamiaa I.A. Metwally, Sally Emad El-Din, Omar Abdelaziz, Iman M. Hamdy, Amr K. Elsamman and Ahmed M. Abdelalim. *Predicting grade of cerebral gliomas using Myo-inositol/Creatine ratio*. Egyptian Journal of Radiology and Nuclear Medicine, vol. 45, no. 1, pages 211–217, 2014. (Cited on page 57.)

[Mocioiu 2016] Victor Mocioiu, Nuno Pedrosa de Barros, Sandra Ortega Martorell, Johannes Slotboom, Urspeter Knecht, Carles Arús, Alfredo Vellido Alcacena and Margarida Julià Sapé. *A machine learning pipeline for supporting differentiation of glioblastomas from single brain metastases*. In ESANN 2016 proceedings: European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning: Bruges (Belgium), 27-29 April 2016, pages 247–252. I6doc. com, 2016. (Cited on pages 172 and 189.)

[Möller-Hartmann 2002] W. Möller-Hartmann, S. Herminghaus, T. Krings, G. Marquardt, H. Lanfermann, U. Pilatus and F. E. Zanella. *Clinical application of proton magnetic resonance spectroscopy in the diagnosis of intracranial mass lesions*, 2002. (Cited on page 58.)

[Moonen 1989] Chrit T W Moonen, Markus Von Kienlin, Peter C M Van Zijl, Jack Cohen, Joe Gillen, Peter Daly and Gerald Wolf. *Comparison of Single-shot Localization Methods (STEAM and PRESS) for in vivo Proton NMR Spectroscopy*. NMR in Biomedicine, vol. 2, no. 5/6, pages 201–208, 1989. (Cited on page 28.)

[Mountford 2010] Carolyn E. Mountford, Peter Stanwell, Alexander Lin, Saadallah Ramadan and Brian Ross. *Neurospectroscopy: The past, present and future*. Chemical Reviews, vol. 110, no. 5, pages 3060–3086, 2010. (Cited on pages 53, 55, 56 and 188.)

[Mueller-Lisse 2007] Ullrich G. Mueller-Lisse and Michael K. Scherr. *Proton MR spectroscopy of the prostate*. European Journal of Radiology, vol. 63, no. 3, pages 351–360, 2007. (Cited on pages 1 and 53.)

[Mulkern 2001] Robert V. Mulkern and Lawrence P. Panych. *Echo planar spectroscopic imaging*. Concepts in Magnetic Resonance, vol. 13, no. 4, pages 213–237, 2001. (Cited on pages 81, 166 and 206.)

[Naser 2016] Radwa Kamel Abdel Naser, Afaf Abdel Kader Hassan, Amr Mohamed Shabana and Nagham Nabil Omar. *Role of magnetic resonance spectroscopy in grading of primary brain tumors*. The Egyptian Journal of Radiology and Nuclear Medicine, vol. 47, no. 2, pages 577–584, 2016. (Cited on page 61.)

[Near 2015] Jamie Near, Richard Edden, C. John Evans, Raphaël Paquin, Ashley Harris and Peter Jezzard. *Frequency and phase drift correction of magnetic resonance spectroscopy data by spectral registration in the time domain*. Magnetic Resonance in Medicine, vol. 73, no. 1, pages 44–50, 2015. (Cited on page 38.)

[Niyazi 2016] Maximilian Niyazi, Michael Brada, Anthony J. Chalmers, Stephanie E. Combs, Sara C. Erridge, Alba Fiorentino, Anca L. Grosu, Frank J. Lagerwaard, Giuseppe Minniti, René Olivier Mirimanoff, Umberto Ricardi, Susan C. Short, Damien C. Weber and Claus Belka. *ESTRO-ACROP guideline "target delineation of glioblastomas"*. Radiotherapy and Oncology, vol. 118, no. 1, pages 35–42, 2016. (Cited on page 172.)

[Noll 2006] D. Noll. Medical Imaging Systems Course Notes. University of Michigan, Ann Arbor, Michigan, USA, 2006. (Cited on page 13.)

[Ocali 1998] Ogan Ocali and Ergin Atalar. *Ultimate intrinsic signal-to-noise ratio in MRI*. Magnetic Resonance in Medicine, vol. 39, no. 3, pages 462–473, 1998. (Cited on page 12.)

[Ortega-Martorell 2012a] Sandra Ortega-Martorell, Paulo J G Lisboa, Alfredo Vellido, Rui V. Simões, Martí Pumarola, Margarida Julià-Sapé and Carles Arús. *Convex Non-Negative Matrix Factorization for Brain Tumor Delimitation from MRSI Data*. PLoS ONE, vol. 7, no. 10, 2012. (Cited on pages 135, 172 and 189.)

[Ortega-Martorell 2012b] Sandra Ortega-Martorell, Paulo J.G. Lisboa, Alfredo Vellido, Margarida Julia-Sape and Carles Arus. *Non-negative Matrix Factorisation methods for the spectral decomposition of MRS data from human brain tumours*. BMC Bioinformatics, vol. 13, no. 1, page 38, 2012. (Cited on pages 44 and 135.)

[Oz 2014] Gülin Oz, Jeffry R Alger, Peter B Barker, Robert Bartha, Alberto Bizzi, Chris Boesch, Patrick J Bolan, Kevin M Brindle, Cristina Cudalbu, Alp Dinçer, Ulrike Dydak, Uzay E Emir, Jens Frahm, Ramón Gilberto González, Stephan Gruber, Rolf Gruetter, Rakesh K Gupta, Arend Heerschap, Anke Henning, Hoby P Hetherington, Franklyn A Howe, Petra S Hüppi, Ralph E Hurd, Kantarci Kantarci, Dennis W J Klomp,

Roland Kreis, Marijn J Kruiskamp, Martin O Leach, Alexander P Lin, Peter R Luijten, Malgorzata Marjańska, Andrew A Maudsley, Dieter J Meyerhoff, Carolyn E Mountford, Sarah J Nelson, M Necmettin Pamir, Jullie W Pan, Andrew C Peet, Harish Poptani, Stefan Posse, Petra J W Pouwels, Eva-Maria Ratai, Brian D Ross, Tom W Scheenen, Christian Schuster, Ian C P Smith, Brian J Soher, Ivan Tkáč, Daniel B Vigneron and Risto A Kauppinen. *Clinical proton MR spectroscopy in central nervous system disorders.* Radiology, vol. 270, no. 3, pages 658–79, 2014. (Cited on pages 80, 97, 148 and 188.)

[Pallud 2010] J. Pallud, P. Varlet, B. Devaux, S. Geha, M. Badoual, C. Deroulers, P. Page, E. Dezamis, C. Daumas-Duport and F. X. Roux. *Diffuse low-grade oligodendrogliomas extend beyond MRI-defined abnormalities*. Neurology, vol. 74, no. 21, pages 1724–1731, 2010. (Cited on pages 62 and 189.)

[Patra 2012] Swarnajyoti Patra and Lorenzo Bruzzone. *A cluster-assumption based batch mode active learning technique*. Pattern Recognition Letters, vol. 33, no. 9, pages 1042–1048, 2012. (Cited on page 120.)

[Pauly 1991] John Pauly, Dwight Nishimura, Albert Macovski and Patrick Le Roux. *Parameter Relations for the Shinnar-Le Roux Selective Excitation Pulse Design Algorithm*. IEEE Transactions on Medical Imaging, vol. 10, no. 1, pages 53–65, 1991. (Cited on page 14.)

[Pearson 2016] Ronald K Pearson, Yrjö Neuvo, Jaakko Astola and Moncef Gabbouj. *Generalized Hampel Filters*. EURASIP Journal on Advances in Signal Processing, 2016. (Cited on page 40.)

[Peden 1993] C. J. Peden, M. A. Rutherford, J. Sargentoni, I. J. Cox, D. J. Bryant and L. M S Dubowitz. *Proton spectroscopy of the neonatal brain following hypoxic-ischaemic injury*. Developmental Medicine & Child Neurology, vol. 35, no. 6, pages 502–510, 1993. (Cited on pages 1 and 53.)

[Pedrosa de Barros 2014] Nuno Pedrosa de Barros, M Jablonski, A Pica, J Starcukova, U Knecht, R Wiest and J Slotboom. *Unifying clinical routine brain tumor MR-Spectoscopy and MR-Image analysis: novel jMRUI plug-ins for brain tumor analysis*. Neuro-Oncology, vol. 16, no. suppl2, pages ii78–ii78, 2014. (Cited on pages 101, 102 and 121.)

[Pedrosa de Barros 2015] Nuno Pedrosa de Barros, V Mocioiu, S Ortega Martorell, U Knecht, C Arús, J Slotboom and M Julià-Sapé. *Highlighting differences between GBM and brain metastasis using a blind source separation method applied to MRSI data*. In ESMRMB 2015, Edinburgh, UK, 2015. (Cited on pages 172 and 189.)

[Pedrosa de Barros 2016a] Nuno Pedrosa de Barros, Urspeter Knecht, Richard McKinley, Jonathan Giezendanner, Roland Wiest and Johannes Slotboom. *Automatic quality assessment of short and long-TE brain tumour MRSI data using novel Spectral Features*. In ISMRM 2016, Singapore, 2016. (Cited on pages 81, 82, 120, 121, 124, 125, 149 and 174.)

[Pedrosa de Barros 2016b] Nuno Pedrosa de Barros, Richard Mckinley, Urspeter Knecht, Roland Wiest and Johannes Slotboom. *Automatic quality control in clinical 1 H MRSI of brain cancer*. NMR in Biomedicine, no. August 2015, 2016. (Cited on pages 81, 82, 83, 120, 121, 122, 124, 125, 142, 149, 164, 165, 170 and 174.)

[Pedrosa de Barros 2017a] Nuno Pedrosa de Barros, Richard Mckinley, Roland Wiest and Johannes Slotboom. *Improving labeling efficiency in automatic quality control of MRSI data*. Magnetic Resonance in Medicine, vol. 00, pages 1–7, 2017. (Cited on page 174.)

[Pedrosa de Barros 2017b] Nuno Pedrosa de Barros and Johannes Slotboom. *Quality management in in vivo proton MRS*. Analytical Biochemistry, vol. 529, pages 98–116, 2017. (Cited on pages 149, 150, 174 and 188.)

[Pelleg 2000] D. Pelleg and A.W. Moore. *X-means: Extending K-means with efficient estimation of the number of clusters*. Proceedings of the Seventeenth International Conference on Machine Learning table of contents, pages 727–734, 2000. (Cited on pages 45, 50, 51, 142, 151 and 203.)

[Pérez-Ruiz 2010] A. Pérez-Ruiz, M. Julià-Sapé, G. Mercadal, I. Olier, C. Majós and C. Arús. *The INTERPRET Decision-Support System version 3.0 for evaluation of Magnetic Resonance Spectroscopy data from human brain tumours and other abnormal brain masses*. BMC Bioinformatics, vol. 11, 2010. (Cited on pages 38 and 75.)

[Philips 2016] Bart W J Philips and Tom W Scheenen. *Methodology of Clinical MRS: Technical Challenges and Solutions*. In Magnetic Resonance Spectroscopy of Degenerative Brain Diseases, pages 31–54. Springer, 2016. (Cited on page 30.)

[Pijnappel 1992] W. W.F. Pijnappel, A. van den Boogaart, R. de Beer and D. van Ormondt. *SVD-based quantification of magnetic resonance signals*. Journal of Magnetic Resonance (1969), vol. 97, no. 1, pages 122–134, 1992. (Cited on page 34.)

[Pirzkall 2004] Andrea Pirzkall, Xiaojuan Li, Joonmi Oh, Susan Chang, Mitchel S. Berger, David A. Larson, Lynn J. Verhey, William P. Dillon and Sarah J. Nelson. *3D MRSI for resected high-grade gliomas before RT: Tumor extent according to metabolic activity in relation to MRI*. International Journal of Radiation Oncology Biology Physics, vol. 59, no. 1, pages 126–137, 2004. (Cited on pages 163 and 177.)

[Podo 1995] F. Podo, W. M M J Bovée, J. de Certaines, D. Leibfritz and J. S. Orr. *Quality assessment in in vivo NMR spectroscopy: I. Introduction, objectives, and activities*. Magnetic Resonance Imaging, vol. 13, no. 1, pages 117–121, 1995. (Cited on pages 68, 70 and 73.)

[Podo 1998] F. Podo, O. Henriksen, W. M M J Bovée, M. O. Leach, D. Leibfritz and J. D. De Certaines. *Absolute metabolite quantification by in vivo nmr spectroscopy: I. Introduction, objectives and activities of a concerted action in biomedical research*. Magnetic Resonance Imaging, vol. 16, no. 9, pages 1085–1092, 1998. (Cited on page 70.)

[Porz 2014] Nicole Porz, Stefan Bauer, Alessia Pica, Philippe Schucht, Jürgen Beck, Rajeev Kumar Verma, Johannes Slotboom, Mauricio Reyes and Roland Wiest. *Multi-modal glioblastoma segmentation: Man versus machine*. PLoS ONE, vol. 9, no. 5, page e96873, 2014. (Cited on pages 141, 142, 148, 150 and 174.)

[Posse 1995] S. Posse, G. Tedeschi, R. Risinger, R. Ogg and D. Le Bihan. *High speed 1H spectroscopic imaging in human brain by echo planar spatial-spectral encoding*. Magnetic Resonance in Medicine, 1995. (Cited on pages 81 and 188.)

[Preul 1996] Mark C. Preul, Zografos Caramanos, D. Louis Collins, Jean-Guy Villemure, Richard Leblanc, Andre Olivier, Ronald Pokrupa and Douglas L. Arnold. *Accurate, noninvasive diagnosis of human brain tumors by using proton magnetic resonance spectroscopy*. Nature Medicine, vol. 2, no. 3, pages 323–325, 1996. (Cited on pages 1, 53 and 148.)

[Preul 1998] MC C Preul, Z Caramanos, R Leblanc, J G Villemure and D L Arnold. *Using pattern analysis of in vivo proton MRSI data to improve the diagnosis and surgical management of patients with brain tumors.* NMR in biomedicine, vol. 11, no. 4-5, pages 192–200, 1998. (Cited on pages 148 and 172.)

[Provencher 1993] S. W. Provencher. *Estimation of metabolite concentrations from localized in vivo proton NMR spectra*. Magnetic Resonance in Medicine, vol. 30, no. 6, pages 672–679, 1993. (Cited on page 87.)

[Purcell 1946] E. M. Purcell, H. C. Torrey and R. V. Pound. *Resonance Absorption by Nucelar Magnetic Moments in a Solid*, 1946. (Cited on page 7.)

[Puts 2012] Nicolaas A J Puts and Richard A E Edden. *In vivo magnetic resonance spectroscopy of GABA: A methodological review*, 2012. (Cited on page 57.)

[Rabi 1938] I Rabi, J. Zacharias, S. Millman and P. Kusch. *A New Method of Measuring Nuclear magnetic Moment*. Physical Review, vol. 53, no. February, page 318, 1938. (Cited on page 7.)

[Raschke 2015] Felix Raschke, Greg A. Fellows, Alan J. Wright and Franklyn A. Howe. *1H 2D MRSI tissue type analysis of gliomas*. Magnetic Resonance in Medicine, vol. 73, no. 4, pages 1381–1389, 2015. (Cited on pages 148, 172 and 189.)

[Ratiney 2004] H. Ratiney, Y. Coenradie, S. Cavassila, D. Van Ormondt and D. Graveron-Demilly. *Time-domain quantitation of 1H short echo-time signals: Background accommodation*. Magnetic Resonance Materials in Physics, Biology and Medicine, vol. 16, no. 6, pages 284–296, 2004. (Cited on pages 42, 142, 149 and 174.)

[Ratiney 2005] H. Ratiney, M. Sdika, Y. Coenradie, S. Cavassila, D. van Ormondt and Danielle Graveron-Demilly. *Time-domain semi-parametric estimation based on a metabolite basis set*. NMR in Biomedicine, vol. 18, no. 1, pages 1–13, 2005. (Cited on pages 42, 56, 86, 87, 100 and 119.)

[Razek 2014] Ahmed Abdel Khalek Abdel Razek, Ahmed Abdalla, Amany Ezzat, Ahmed Megahed and Tarek Barakat. *Minimal hepatic encephalopathy in children with liver cirrhosis: diffusion-weighted MR imaging and proton MR spectroscopy of the brain*. Neuroradiology, vol. 56, no. 10, pages 885–891, 2014. (Cited on pages 1 and 53.)

[Ricci 2007] Renata Ricci, A. Bacci, V. Tugnoli, S. Battaglia, M. Maffei, R. Agati and M. Leonardi. *Metabolic findings on 3T 1H-MR spectroscopy in peritumoral brain edema*. American Journal of Neuroradiology, vol. 28, no. 7, pages 1287–1291, 2007. (Cited on pages 151, 152, 153, 179 and 189.)

[Rios Velazquez 2015] Emmanuel Rios Velazquez, Raphael Meier, William D Dunn Jr, Brian Alexander, Roland Wiest, Stefan Bauer, David A Gutman, Mauricio Reyes and Hugo J W L Aerts. *Fully automatic GBM segmentation in the TCGA-GBM dataset: Prognosis and correlation with VASARI features*. Scientific Reports, vol. 5, page 16822, nov 2015. (Cited on pages 163 and 167.)

[Rock 2004] Jack P. Rock, Lisa Scarpace, David Hearshen, Jorge Gutierrez, James L. Fisher, Mark Rosenblum, Tom Mikkelsen, Joachim Baehring, Bruce E. Pollock, Douglas Kondziolka, Michelle S. Bradbury, Philip H. Gutin, Jeffrey Weinberg and Raymond Sawaya. *Associations among Magnetic Resonance Spectroscopy, Apparent Diffusion Coefficients, and Image-guided Histopathology with Special Attention to Radiation Necrosis*. Neurosurgery, vol. 54, no. 5, pages 1111–1119, 2004. (Cited on page 62.)

[Rule 2006] G S Rule and T K Hitchens. Fundamentals of Protein NMR Spectroscopy, volume 5. 2006. (Cited on pages 23 and 25.)

[Sabati 2015] Mohammad Sabati, Sulaiman Sheriff, Meng Gu, Juan Wei, Henry Zhu, Peter B. Barker, Daniel M. Spielman, Jeffry R. Alger and Andrew A. Maudsley. *Multivendor implementation and comparison of volumetric whole-brain echo-planar MR spectroscopic imaging*. In Magnetic Resonance in Medicine, volume 74, pages 1209–1220, 2015. (Cited on page 188.)

[Sajda 2004] Paul Sajda, Shuyan Du, Truman R. Brown, Radka Stoyanova, Dikoma C. Shungu, Xiangling Mao and Lucas C. Parra. *Nonnegative matrix factorization for rapid recovery of constituent spectra in magnetic resonance chemical shift imaging of the brain*. IEEE Transactions on Medical Imaging, vol. 23, no. 12, pages 1453–1465, 2004. (Cited on pages 172 and 189.)

[Scheenen 2008] Tom W.J. Scheenen, Dennis W.J. Klomp, Jannie P. Wijnen and Arend Heerschap. *Short echo time 1H-MRSI of the human brain at 3T with minimal chemical shift displacement errors using adiabatic refocusing pulses*. Magnetic Resonance in Medicine, vol. 59, no. 1, pages 1–6, 2008. (Cited on page 30.)

[Schucht 2014] Philippe Schucht, Sonja Knittel, Johannes Slotboom, Kathleen Seidel, Michael Murek, Astrid Jilch, Andreas Raabe and Jürgen Beck. *5-ALA complete resections go beyond MR contrast enhancement: Shift corrected volumetric analysis of the extent of resection in surgery for glioblastoma*. In Acta Neurochirurgica, volume 156, pages 305–312, 2014. (Cited on page 172.)

[Segal 2004] Mark R Segal. *Machine Learning Benchmarks and Random Forest Regression*. Biostatistics, no. MAY 2003, pages 1–14, 2004. (Cited on pages 48 and 106.)

[Server 2010] Andrès Server, Roger Josefsen, Bettina Kulle, Jan Maehlen, Till Schellhorn, Øystein Gadmar, Theresa Kumar, Monika Haakonsen, Carl W Langberg and Per H Nakstad. *Proton magnetic resonance spectroscopy in the distinction of high-grade cerebral gliomas from single metastatic brain tumors.* Acta radiologica (Stockholm, Sweden : 1987), vol. 51, no. 3, pages 316–25, 2010. (Cited on pages 60, 61 and 203.)

[Shepp 1974] L a Shepp and B F Logan. *The Fourier reconstruction of a head section*, 1974. (Cited on page 78.)

[Silva 2004] C. Silva and B. Ribeiro. *Margin-based active learning and background knowledge in text mining*. Fourth International Conference on Hybrid Intelligent Systems (HIS'04), pages 0–5, 2004. (Cited on page 120.)

[Silver 2016] David Silver, Aja Huang, Chris J. Maddison, Arthur Guez, Laurent Sifre, George van den Driessche, Julian Schrittwieser, Ioannis Antonoglou, Veda Panneershelvam, Marc Lanctot, Sander Dieleman, Dominik Grewe, John Nham, Nal Kalchbrenner, Ilya Sutskever, Timothy Lillicrap, Madeleine Leach, Koray Kavukcuoglu, Thore Graepel and Demis Hassabis. *Mastering the game of Go with deep neural networks and tree search*. Nature, vol. 529, no. 7587, pages 484–489, 2016. (Cited on page 43.)

[Silver 2017] David Silver, Julian Schrittwieser, Karen Simonyan, Ioannis Antonoglou, Aja Huang, Arthur Guez, Thomas Hubert, Lucas Baker, Matthew Lai, Adrian Bolton, Yutian Chen, Timothy Lillicrap, Fan Hui, Laurent Sifre, George van den Driessche, Thore Graepel and Demis Hassabis. *Mastering the game of Go without human knowledge*. Nature, vol. 550, no. 7676, pages 354–359, 2017. (Cited on page 43.)

[Slotboom 1991] J Slotboom, A F Mehlkopf and WMMJ Bovee. *A single-shot localization pulse sequence suited for coils with inhomogeneous RF fields using adiabatic slice-selective RF pulses*. Journal of Magnetic Resonance (1969), vol. 95, no. 2, pages 396–404, 1991. (Cited on pages 23, 30, 74 and 150.)

[Slotboom 1993] J. Slotboom. Selective RF Pulses in NMR and Their Effect on Coupled and Uncoupled Spin Systems. TU Delft, Delft University of Technology, Delft, 1993. (Cited on pages 13, 14, 15, 74 and 93.)

[Slotboom 1994a] J. Slotboom, A.F. Mehlkopf and W.M.M.J. Bovee. *The Bloch Equations for an AB System and the Design of Spin-State-Selective RF Pulses for Coupled Spin Systems*. Journal of Magnetic Resonance, Series A, vol. 106, no. 2, pages 159–167, feb 1994. (Cited on page 24.)

[Slotboom 1994b] J. Slotboom, A.F. Mehlkopf and W.M.M.J. Bovee. *The Effects of Frequency-Selective RF Pulses on J-Coupled Spin- Systems*, 1994. (Cited on pages 30, 90, 92 and 150.)

[Slotboom 1998] Johannes Slotboom, Chris Boesch and Roland Kreis. *Versatile frequency domain fitting using time domain models and prior knowledge*. Magnetic Resonance in Medicine, vol. 39, no. 6, pages 899–911, 1998. (Cited on pages 42 and 87.)

[Slotboom 2009] J Slotboom, a Nirkko, C Brekenfeld and D van Ormondt. *Reliability testing of in vivo magnetic resonance spectroscopy (MRS) signals and signal artifact reduction by order statistic filtering*. Measurement Science and Technology, vol. 20, no. 10, page 104030, 2009. (Cited on pages 79, 86, 87, 93, 100 and 119.)

[Slotboom 2011] Johannes Slotboom, Dirk van Ormondt, Olivier Scheidegger, C Brekenfeld, Roland Wiest, Gerhard Schroth and Danielle Graveron-Demilly. *Experimental Errors as Alternative to the Theoretical Cramér-Rao Minimum Variance Bounds in MRS : ER-ARSOS - Error Estimation by Multiple Quantification of Recombined ARSOS-filtered Output Signals*. Proceedings of the 19th Scientific Meeting International Society for Magnetic Resonance in Medicine, vol. 104030, no. 2009, page 104035, 2011. (Cited on pages 86, 87, 100 and 119.)

[Son 2001] B.-C. Son, M.-C. Kim, B.-G. Choi, E.-N. Kim, H.-M. Baik, B.-Y. Choe, S Naruse and J.-K. Kang. *Proton Magnetic Resonance Chemical Shift Imaging (1H CSI)-Directed Stereotactic Biopsy*. Acta Neurochirurgica, vol. 143, no. 1, pages 45–50, jan 2001. (Cited on page 167.)

[Stadlbauer 2004] Andreas Stadlbauer, Ewald Moser, Stephan Gruber, Rolf Buslei, Christopher Nimsky, Rudolf Fahlbusch and Oliver Ganslandt. *Improved delineation of brain tumors: An automated method for segmentation based on pathologic changes of 1H-MRSI metabolites in gliomas*. NeuroImage, vol. 23, no. 2, pages 454–461, 2004. (Cited on pages 62, 143, 147, 151, 152, 167, 172 and 183.)

[Strasser 2012] Bernhard Strasser. *Increasing the signal to noise ratio in MR spectroscopic imaging by coil combination at 7T*. PhD thesis, uniwien, 2012. (Cited on page 95.)

[Su 2008] Yuzhuo Su, Sunitha B. Thakur, Karimi Sasan, Shuyan Du, Paul Sajda, Wei Huang and Lucas C. Parra. *Spectrum separation resolves partial-volume effect of MRSI as demonstrated on brain tumor scans*. NMR in Biomedicine, vol. 21, no. 10, pages 1030–1042, dec 2008. (Cited on pages 172 and 189.)

[Suppiah 2013] S. Suppiah, K. Rahmat, M. N. Mohd-Shah, C. A. Azlan, L. K. Tan, Y. F A Aziz, A. Vijayananthan, A. L. Wui and C. H. Yip. *Improved diagnostic accuracy in differentiating malignant and benign lesions using single-voxel proton MRS of the breast at 3 T MRI*. Clinical Radiology, vol. 68, no. 9, 2013. (Cited on pages 1 and 53.)

[Takano 2001] T Takano, J H Lin, G Arcuino, Q Gao, J Yang and M Nedergaard. *Glutamate release promotes growth of malignant gliomas.* Nature medicine, vol. 7, no. 9, pages 1010–5, 2001. (Cited on pages 55, 165 and 189.)

[Tate 2006] Anne R. Tate, Joshua Underwood, Dionisio M. Acosta, Margarida Julià-Sapé, Carles Majós, Ángel Moreno-Torres, Franklyn A. Howe, Marinette van der Graaf, Virginie Lefournier, Mary M. Murphy, Alison Loosemore, Christophe Ladroue, Pieter Wesseling, Jean Luc Bosson, Miguel E. Cabañas, Arjan W. Simonetti, Witold Gajewicz, Jorge Calvar, Antoni Capdevilla, Peter R. Wilkins, B. Anthony Bell, Chantal Rémy, Arend Heerschap, Des Watson, John R. Griffiths and Carles Arús. *Development of a decision support system for diagnosis and grading of brain tumours using in vivo magnetic resonance single voxel spectra*. NMR in Biomedicine, vol. 19, no. 4, pages 411–434, 2006. (Cited on pages 75 and 100.)

[Tayari 2017] Nassim Tayari, Arend Heerschap, Tom W.J. Scheenen and Thiele Kobus. *In vivo MR spectroscopic imaging of the prostate, from application to interpretation*. Analytical Biochemistry, 2017. (Cited on pages 1 and 53.)

[Tuia 2009] Devis Tuia, Frdric Ratle, Fabio Pacifici, Mikhail F. Kanevski and William J. Emery. *Active learning methods for remote sensing image classification*. IEEE Transactions on Geoscience and Remote Sensing, vol. 47, no. 7, pages 2218–2232, 2009. (Cited on page 120.)

[Tukey 1962] John W. Tukey. *The Future of Data Analysis*. The Annals of Mathematical Statistics, vol. 33, no. 1, pages 1–67, 1962. (Cited on page 151.)

[van der Graaf 2008] Marinette van der Graaf, Margarida Julià-Sapé, Franklyn A. Howe, Anne Ziegler, Carles Majós, Ángel Moreno-Torres, Mark Rijpkema, Dionisio Acosta, Kirstie S. Opstad, Yvonne M. van der Meulen, Carles Arús and Arend Heerschap. *MRS quality assessment in a multicentre study on MRS-based classification of brain tumours*. NMR in Biomedicine, vol. 21, no. 2, pages 148–158, feb 2008. (Cited on pages 68, 75, 81, 82 and 100.)

[Van Der Maaten 2008] L J P Van Der Maaten and G E Hinton. *Visualizing high-dimensional data using t-sne*. Journal of Machine Learning Research, vol. 9, pages 2579–2605, 2008. (Cited on pages 155, 161 and 162.)

[Van Vaals 1990] J. J. Van Vaals and A. H. Bergman. *Optimization of eddy-current compensation*. Journal of Magnetic Resonance (1969), vol. 90, no. 1, pages 52–70, 1990. (Cited on page 88.)

[Vigneron 2001] D Vigneron, a Bollen, M McDermott, L Wald, M Day, S Moyher-Noworolski, R Henry, S Chang, M Berger, W Dillon and S Nelson. *Three-dimensional magnetic resonance spectroscopic imaging of histologically confirmed brain tumors*. Magn Reson Imaging, vol. 19, no. 1, pages 89–101, 2001. (Cited on pages 151, 152 and 167.)

[Webb 2003] Andrew Webb and George C. Kagadis. *Introduction to Biomedical Imaging*. Medical Physics, vol. 30, no. 8, page 2267, 2003. (Cited on page 11.)

[Weller 2014] Michael Weller, Martin van den Bent, Kirsten Hopkins, Jörg C. Tonn, Roger Stupp, Andrea Falini, Elizabeth Cohen-Jonathan-Moyal, Didier Frappaz, Roger Henriksson, Carmen Balana, Olivier Chinot, Zvi Ram, Guido Reifenberger, Riccardo Soffietti and Wolfgang Wick. *EANO guideline for the diagnosis and treatment of anaplastic gliomas and glioblastoma*. The Lancet Oncology, vol. 15, no. 9, pages 395–403, 2014. (Cited on page 172.)

[Weybright 2005] Patrick Weybright, Pia C. Sundgren, Pavel Maly, Diana Gomez Hassan, Bin Nan, Suzan Rohrer and Larry Junck. *Differentiation between brain tumor recurrence and radiation injury using MR spectroscopy*. American Journal of Roentgenology, vol. 185, no. 6, pages 1471–1476, 2005. (Cited on page 62.)

[Wijnen 2012] J. P. Wijnen, A. J S Idema, M. Stawicki, M. W. Lagemaat, P. Wesseling, A. J. Wright, T. W J Scheenen and A. Heerschap. *Quantitative short echo time 1H MRSI of the peripheral edematous region of human brain tumors in the differentiation between glioblastoma, metastasis, and meningioma*. Journal of Magnetic Resonance Imaging, vol. 36, no. 5, pages 1072–1082, 2012. (Cited on pages 61, 140 and 203.)

[Woo 2007] Dong Cheol Woo, Bum Soo Kim, So Lyung Jung, Hyo Jin Park, Hyang Shuk Rhim, Geon Ho Jahng and Bo Young Choe. *Development of a cone-shape phantom for multi-voxel MR spectroscopy*. Journal of Neuroscience Methods, vol. 162, no. 1-2, pages 101–107, 2007. (Cited on page 77.)

[Wright 2008] Alan J. Wright, Carles Arús, Jannie P. Wijnen, Ángel Moreno-Torres, John R. Griffiths, Bernardo Celda and Franklyn A. Howe. *Automated quality control protocol for MR spectra of brain tumors*. Magnetic Resonance in Medicine, vol. 59, no. 6, pages 1274–1281, 2008. (Cited on pages 81, 82 and 100.)

[Wright 2013] Alan J Wright, Thiele Kobus, Kirsten M Selnaes, Ingrid S Gribbestad, Elizabeth Weiland, Tom W J Scheenen and Arend Heerschap. *Quality control of prostate 1 H MRSI data*. NMR in Biomedicine, vol. 26, no. 2, pages 193–203, 2013. (Cited on pages 81, 82, 100, 108, 113 and 120.)

[Yamahara 2010] Takahiro Yamahara, Yoshihiro Numa, Tetsuya Oishi, Takuya Kawaguchi, Toshitaka Seno, Akio Asai and Keiji Kawamoto. *Morphological and flow cytometric analysis of cell infiltration in glioblastoma: A comparison of autopsy brain and neuroimaging*. Brain Tumor Pathology, vol. 27, no. 2, pages 81–87, 2010. (Cited on pages 62 and 172.)

[Yan 2002] H. Yan. *Signal processing for magnetic resonance imaging and spectroscopy*. CRC Press, 2002. (Cited on page 100.)

[Yang 2015] Guang Yang, Felix Raschke, Thomas R. Barrick and Franklyn a. Howe. *Manifold Learning in MR spectroscopy using nonlinear dimensionality reduction and unsupervised clustering*. Magnetic Resonance in Medicine, vol. 74, no. 3, pages 868–878, 2015. (Cited on pages 148 and 172.)

[Zeng 2011] QingShi Zeng, HePeng Liu, Kai Zhang, ChuanFu Li and GengYin Zhou. *Noninvasive evaluation of cerebral glioma grade by using multivoxel 3D proton MR spectroscopy*. Magnetic Resonance Imaging, vol. 29, no. 1, pages 25–31, 2011. (Cited on pages 147, 151, 152, 158 and 167.)

[Zhu 2008] M Zhu, A Bashir, J J Ackerman and D A Yablonskiy. *Improved calibration technique for in vivo proton MRS thermometry for brain temperature measurement*. Magnetic resonance in medicine : official journal of the Society of Magnetic Resonance in Medicine / Society of Magnetic Resonance in Medicine, vol. 60, no. 3, pages 536–541, 2008. (Cited on page 38.)

[Zhu 2011] He Zhu and Peter B Barker. *MR spectroscopy and spectroscopic imaging of the brain.* Methods in molecular biology (Clifton, N.J.), vol. 711, pages 203–226, 2011. (Cited on page 30.)

[Zimmerman 2011] Robert A. Zimmerman. *Neuroimaging of inherited metabolic disorders producing seizures*, 2011. (Cited on pages 1 and 53.)

[Zou 2011] Q. G. Zou, H. B. Xu, F. Liu, W. Guo, X. C. Kong and Y. Wu. *In the assessment of supratentorial glioma grade: The combined role of multivoxel proton MR spectroscopy and diffusion tensor imaging*. Clinical Radiology, vol. 66, no. 10, pages 953–960, 2011. (Cited on pages 147 and 151.)